The International Handbook of Physics Education Research: Special Topics

Citation: *The International Handbook of Physics Education Research: Special Topics*, edited by Mehmet Fatih Taşar and Paula R. L. Heron [AIP Publishing (online), Melville, New York, 2023], available at: https://doi.org/10.1063/9780735425514

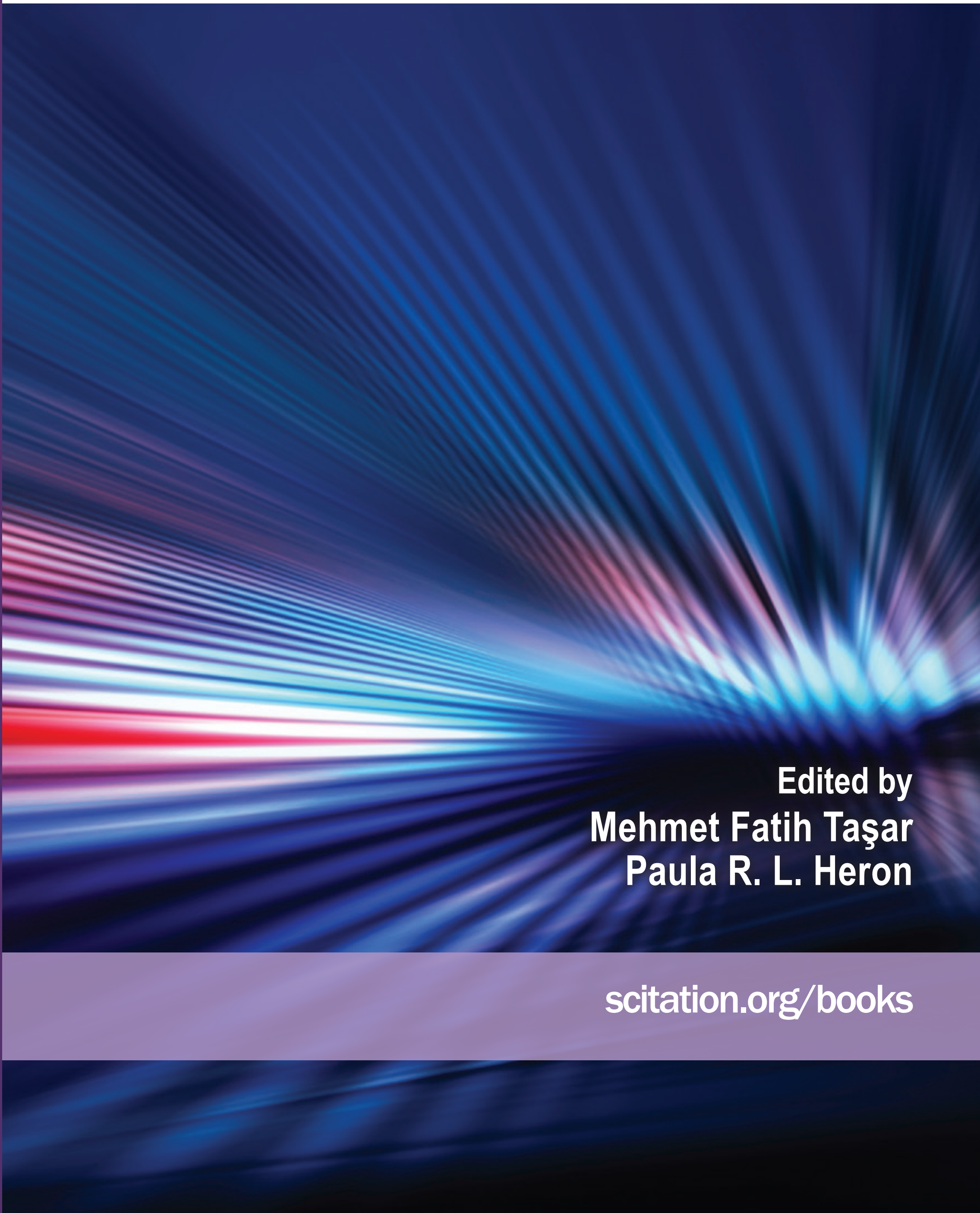
View online: https://doi.org/10.1063/9780735425514

View Table of Contents: https://aip.scitation.org/doi/book/10.1063/9780735425514

Published by AIP Publishing

Professional

The International Handbook of Physics Education Research: Special Topics



Edited by
Mehmet Fatih Taşar
Paula R. L. Heron

AAPT PHYSICS EDUCATION®

AIP Publishing

scitation.org/books

The International Handbook of Physics Education Research: Special Topics

The International Handbook of Physics Education Research: Special Topics

Edited by

Mehmet Fatih Taşar
Georgia State University, Atlanta, GA, USA

Paula R. L. Heron
University of Washington, Seattle, WA, USA

AIP Publishing Books

A publication of AIP Publishing

Melville, New York

First edition, published 2023.

Library of Congress Control Number: 2022950113

ISBN: 978-0-7354-2548-4 (Softcover)
ISBN: 978-0-7354-2551-4 (Online)
ISBN: 978-0-7354-2549-1 (ePub)
ISBN: 978-0-7354-2550-7 (ePDF)

Set in 10/13pt MinionPro by Nova Techset Private Limited, Bengaluru & Chennai, India, and bound by The Sheridan Group, USA

Published by AIP Publishing
1305 Walt Whitman Road, Suite 110, Melville, NY 11747-4300, USA

PREFACE

As a young science education researcher at Penn State in the mid-1990s with a background in pure physics (condensed matter—high temperature superconductivity), I was totally perplexed and worried about the future of my doctoral studies. I was new to the field and did not know where to turn for research ideas and, if I found one, how to validate the rationale. When I progressed further into my coursework, we were introduced to the handbooks that existed at the time: the *Handbook of Research on Science Teaching and Learning* (Gabel, 1994) and the *International Handbook of Science Education* (*IHSE*) (Fraser and Tobin, 1998). I was amazed by the breadth and depth of the chapters written by experts on each topic deemed to be of concern and be attractive to science education researchers. The sections and chapters outline the major research areas and a respective synthesis of research. Later, other handbooks followed, both those covering science education broadly [e.g., *The Handbook of Research on Science Education* (*HRSE*) (Abell and Lederman, 2007 and 2014) and *The Second International Handbook of Science Education* (Fraser, Tobin, McRobbie, 2012)], and others that covered specific topics in depth [e.g., *The International Handbook of Research on Conceptual Change* (Vosniadou, 2008) and *The International Handbook of Research in History, Philosophy and Science Teaching* (Matthews, 2014)].

While *IHSE* did not have chapters on specific science content areas (i.e., physics, chemistry, and biology), *HRSE* included such chapters in its "Science Teaching" section. Reinders Duit has been compiling a bibliography of science education for some decades and reporting percentages of published studies for each content area. In *HRSE*'s "Teaching Physics" section, the authors (Duit *et al.*) reported that "… according to the bibliography on constructivist-oriented research on teaching and learning science by Duit (2009), about 53% of the studies documented were carried out in the domain of physics, 18% in the domain of biology, and 28% in the domain of chemistry." To cut a long story short, having known the obvious advantages of handbooks and the fact that the field of physics education research has a high number of published studies, I envisioned editing a handbook dedicated to physics education research (PER). But the timing was also crucial, and I was thinking that it was ripening already in 2018.

Scholarly contributions to PER come from two types of researchers: those coming from a background of physics teaching and science education research studies in a college/faculty of education—and those coming from a background of college/university level physics teaching and PER in a department of physics. I mostly represent the former and I thought I needed a co-editor from the background of the latter. Although I had my doctoral degree from an American university, since 2001 I had resided in Türkiye and had become very active in European science education and physics education circles. Moreover, as an educator, I valued and practiced actions favoring inclusion and diversity throughout my professional life. I have cherished international collaborations and connections from around the world and always kept in mind that education and educational research is about enhancing human capacities.

With all these thoughts in my mind, and given the fact that American contributions to PER have been immense, I wanted to have a colleague from the USA co-edit the handbook with me. Thus, I decided to approach Paula Heron, who has a Ph.D. in Physics and is in a Physics department. We had already known each other for quite some time, and I very much respected and admired her contributions to PER, just like everyone else did in our field. I emailed Paula in early March of 2019. She was a keynote speaker at the GIREP conference that was going to be held in Budapest in early July of that year. Paula carefully considered my invitation to co-edit *The International Handbook of Physics Education Research* (*IHPER*) and within a few days responded, as she promised, with a positive answer. Committing oneself to a long-term project like this is indeed courageous and for that reason I am forever grateful to Paula for teaming up with me in this extremely important endeavor, the value of which I am confident will be appreciated in the years to come.

Paula and I could be a successful team of co-editors once we set clear goals and plans, show strong leadership to achieve those goals, fulfill our own tasks, and also help each other communicate openly, resolve emerging conflicts constructively, and feel that each one of us is directly contributing to the handbook's success. All of these became true over the course of the creation of *IHPER*. I am forever grateful to Paula for being such a wonderful colleague and co-editor.

In Budapest, we met and talked about some of the details of the project. Also, since many PER people were already there, it was a precious opportunity for us to open the project to potential contributors, collect their ideas, and seek ways to involve them in *IHPER*. The next steps were to form a structure and organization for *IHPER* and find a publisher. Later, we formed an international advisory board to share the idea of *IHPER* and their views about the draft structure and organization. As a result, we received much praise and positive feedback. Among our efforts to find a publisher for *IHPER*, we finally contacted the AAPT Committee on Publications, who had an agreement with the American Institute of Physics (AIP) to publish books. AIP Publishing reviewed our proposal for *IHPER* and in July 2020, we signed a contract.

For *IHPER* to deserve the "international" character in its name, we wanted to include colleagues with extensive experience in PER from around the world. Another aspect was to have diverse teams of co-authors, such as relatively new and relatively experienced ones, and ones from different countries (or better, whenever possible, from different continents). In addition, we wanted to share not only the responsibility and workload but also the joy and pride of creating *IHPER* with respected PER colleagues. Therefore, we decided to have section editors collaborate with us in identifying chapter authors and tracking progress. To a large degree, our scheme worked.

It is important to note that the development of the *IHPER* took place during a time of tremendous upheaval and uncertainty. The global Covid-19 pandemic presented editors, authors, and reviewers with unanticipated challenges in maintaining high standards while meeting publisher deadlines. While schools and businesses were closed and lockdowns were ordered, we all experienced difficult times. But, the work had to go on. Paula and I held weekly online video meetings and had meetings with section editors. It was a big challenge to organize it since we were spread out around the world. Nevertheless, things worked out well. Afterwards, we communicated frequently with the section editors to respond to their questions, to provide initial editorial reviews for submitted first drafts of chapters, to recruit reviewers, and anything else that came along.

Initially, I was in Türkiye and Paula was in the U.S. During the last year, we switched continents. She came to Europe for a sabbatical, while I moved to the U.S. We still had several hours of time difference, but it did not stop us from working together in accordance with our determination to successfully complete *IHPER*. That was our great responsibility to so many who vested trust in us and have been devoting their time and efforts as section editors, authors, and reviewers with diligence, motivation, and ambition.

Now, we have the final manuscript, which consists of three volumes organized into 12 sections, with a total of 69 chapters. Nineteen section editors and 170 authors contributed and benefited from the expertise of many external reviewers. Section editors, contributors and reviewers represent countries from Europe, the Americas, Asia, Africa, the Middle East, and Australia. As general editors, we are greatly thankful to all.

M. Fatih Taşar
Georgia State University
Atlanta, GA, USA

Coming from a physics background with a Ph.D. in Theoretical Physics, and entering the field of PER as a postdoc, handbooks were not a significant part of my early professional development. In the mid-1990s, when I joined the Physics Education Group at the University of Washington (then under the leadership of Lillian C. McDermott), the field seemed small enough that everyone knew everyone else (at least in the U.S.) and you could pick up the phone or send an email to inquire about what they were up to. The literature was relatively sparse, especially concerning university-level teaching, and most researchers were intimately familiar with a small set of seminal papers. Since then, the field has grown enormously and I have come to appreciate the value of review articles, such as those found in handbooks. In my role as an Associate Editor of Physical Review—PER, it has frequently been the case that I have needed a quick overview of a particular area of research. This, more than anything else, convinced me that a handbook for PER would be an invaluable resource for our field. I am grateful that Fatih approached me about this project, which seemed ambitious at first, but has grown into something even bigger than I think either of us imagined. I am also deeply appreciative of all of the effort that has gone into it, especially by the section editors, without whom the project would not have been possible.

Paula R. L. Heron
University of Washington
Seattle, WA, USA

ACKNOWLEDGMENTS

This handbook would not be possible without the support, suggestions, critiques, and contributions of a large number of people. First and foremost, we are grateful for the work of the section editors who recruited and supported authors, reviewed chapters, and wrote section overviews: Marisa Michelini, Italy; Shulamit Kapon, Israel; Olivia Levrini, Italy; Edit Yerushalmi, Israel; Bat Sheva Eyon, Israel; Sarantos Psycharis, Greece; Eugenia Etkina, USA; Eric Brewe, USA; Eilish McLoughlin, Ireland; Feral Ogan Bekiroglu, Turkiye; Geraldine Cochran, USA; Don Metz, Canada; Peter Heering, Germany; Cibelle Celestino Silva, Brazil; Marika Kapanadze, Georgia; Gabriela Jonas-Ahrend, Germany; Gesche Pospiech, Germany; and David Meltzer, USA. The results are a credit to their efforts. Early in the process we received invaluable suggestions from a small group of advisors: Richard Gunstone, Australia; David Hestenes, USA; the late Norman G. Lederman, USA; Marcia Linn, USA; Robin Millar, UK; E. F. (Joe) Redish, USA; the late Laurence Viennot, France; and the late Lillian C. McDermott, USA. Blane Baker (then AAPT Secretary) and Laura McCullough (then Chair of the AAPT Books Advisory Committee) were instrumental in securing the sponsorship of the AAPT. Martine Felton, Michael Lynch, and Dina Rabie at AIP Publishing worked with us to prepare the manuscript.

We are also grateful for the review efforts and invaluable feedback provided for improving the manuscripts in this volume by the following colleagues:

Section I

Alexis Knaub, American Association of Physics Teachers, USA

Ramón Barthelemy, University of Utah, USA

Diana Sachmpazidi, University of Maryland, USA

Jacquelyn J. Chini, University of Central Florida, USA

Erin M. Scanlon, University of Connecticut

Sonja Cwik, University of Pittsburgh, USA

Chandralekha Singh, University of Pittsburgh, USA

Section II

Peter Heering, Institute of Mathematical, Scientific and Technical Literacy, Europa-

Universitaet Flensburg, Germany

Donald J. Metz, University of Winnipeg, Canada

Cibelle Silvestra Silva, Instituto de Fisica de Sao Carlos, Universidade de Sao Paulo, Brazil

Section III

Hendrik Haertig, University of Duisburg-Essen, Germany

Peter Labudde, University of Applied Sciences Northwestern Switzerland, School of Teacher Education, Switzerland

Yaron Lehavi, The David Yellin Academic College of Education, Israel

Charles Ruggieri, Department of Physics and Astronomy, School of Arts and Sciences. Rutgers, The State University of New Jersey, USA

Section IV

Marco Caballero, Michigan State University, USA

Mieke de Cock, KU Leuven, Belgium

Tor Ole Odden, University of Oslo, Norway

Section V

Andrew F. Heckler, The Ohio State University, USA

Steven Pollock, University of Colorado, Boulder, USA

Eric Brewe, Drexel University, USA

Robert Talbot, University of Colorado, Denver, USA

Eugenia Etkina, Rutgers University, USA

Andrew Boudreaux, Western Washington University, USA

Trevor Smith, Rowan University, USA

CONTRIBUTORS

Jaume Ametller
Department of Specific Didactics, University of Girona, Spain

Hillary Diane Andales
MIT Edgerton Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Katrin Bölsterli Bardy
University of Teacher Education Lucerne, CH-6003 Lucerne, Switzerland

Ramón S. Barthelemy
Department of Physics and Astronomy, University of Utah, Salt Lake City, Utah 84112, USA

Jennifer Blue
Department of Physics, Miami University, Oxford, Ohio 45056, USA

Elizabeth Mary Cavicchi
MIT Edgerton Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Yongwook Cheong
Department of Physics Education, Gyeongsang National University, Gyeongsangnam-do, South Korea

Jacquelyn J. Chini
Department of Physics, College of Sciences, University of Central Florida, Orlando, Florida 32816, USA

Sonja Cwik
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

Thaís Cyrino de Mello Forato
Federal University of Sao Paolo—Universidade Federal de São Paulo, Brazil

Isabelle Priscila Carneiro de Lima
Federal Institute of Bahia, Bahia, Brazil

Lin Ding
Department of Teaching and Learning, The Ohio State University, Columbus, Ohio 43210, USA

Elias Euler
Colorado School of Mines, Golden, Colorado 80401, USA

Gabriela Kaiana Ferreira
Physics Department, Federal University of Santa Catarina, Brazil

Johannes Grebe-Ellis
School of Mathematics and Natural Sciences, Bergische Universität Wuppertal, 42119 Wuppertal, Germany

Bor Gregorcic
Department of Physics and Astronomy, Uppsala University, 751 20 Uppsala, Sweden

Sascha Grusche
University Library, Technical University of Munich, D-80333 München, Germany

Andreia Guerra
Graduate and Research Department, CEFET/RJ, Rio de Janeiro, Brazil

Jenaro Guisasola
Department of Applied Physics, School of Engineering Gipuzkoa, University of the Basque, Country (UPV/EHU), Donostia, Spain

Ivã Gurgel
Physics Institute, University of São Paulo, São Paulo, Brazil

Jesper Haglund
Department of Engineering and Physics, Karlstad University, 651 88 Karlstad, Sweden

David Hammer
Department of Education and Department of Physics and Astronomy, Tufts University, Medford, Massachusetts 02155, USA

John Hansen
Department of Physics and Astronomy, West Virginia University, Morgantown, West Virginia 26506, USA

Danielle Boyd Harlow
Gevirtz Graduate School of Education, University of California, Santa Barbara, California 93106-9490, USA

Yuze He

Beijing Normal University, Beijing, China

Larry V. Hedges

Department of Statistics and Data Science, Northwestern University, Evanston, Illinois 60208, USA

Lana Ivanjek

TU Dresden, Haeckelstraße 3, 01069 Dresden, Germany

Hunkoog Jho

Graduate School of Education, Dankook University, Gyeonggi-do 16890, South Korea

Gabriela Jonas-Ahrend

Department of Technology Didactics, Paderborn University, Paderborn, Germany

Fadeel Joubran

Department of Physics, The Academic Arab College for Education, Haifa 3442600, Israel

Calvin S. Kalman

Department of Physics, Concordia University, Montreal, Quebec, Canada

Marika Kapanadze

Ilia State University, Tbilisi, Georgia

Ricardo Avelar Sotomaior Karam

Department of Science Education, University of Copenhagen, Copenhagen, Denmark

Minchul Kim

Department of Physics Education, Kongju National University, Chungcheongnam-do, South Korea

Alexis V. Knaubn

American Association of Physics Teachers, USA

Xiying Li

Key Laboratory of Modern Teaching Technology, Ministry of Education, 710062, People's Republic of China

Alexander Mazzolini

Department of Physics and Astronomy, Swinburne University of Technology, Melbourne, Australia

David E. Meltzer

College of Integrative Sciences and Arts, Arizona State University, Mesa, Arizona 85212, USA

Riley S. Moeykens

MIT Edgerton Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Valerie K. Otero

School of Education, University of Colorado at Boulder, Boulder, Colorado 80309-0249, USA

Oliver Passon

School of Mathematics and Natural Sciences, Bergische Universität Wuppertal, 42119 Wuppertal, Germany

Xiaomei Ping

Beijing Normal University, Beijing, China

Maja Planinic

Department of Physics, Faculty of Science, University of Zagreb, Bijenicka c. 32, HR-10000 Zagreb, Croatia

Gesche Pospiech

Faculty of Physics, TU Dresden, 01062 Dresden, Germany

Diana Sachmpazidi

Department of Physics, College of Computer, Mathematical, and Natural Sciences, University of Maryland, College Park, Maryland 20742, USA

Paulo Sarriugarte

Department of Applied Physics, School of Engineering Bilbao, University of the Basque Country (UPV/EHU), Bilbao, Spain

Erin M. Scanlon

Department of Physics, University of Connecticut—Avery Point, Groton, Connecticut 06340, USA

Roland M. Schulz

CIRCE SFU, 1558 Roxbury Rd, North Vancouver, Vancouver, British Columbia, V7G 1X7, Canada

Chandralekha Singh

Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

Josip Slisko

Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, Puebla, Mexico

Lorena Solvang

Department of Engineering and Physics, Karlstad University, 651 88 Karlstad, Sweden

Jinwoong Song

Department of Physics Education, Seoul National University, Seoul, South Korea

Richard Staley

Department of History and Philosophy of Science, University of Cambridge, Cambridge, CB2 3RH, United Kingdom

Department of Science Education, University of Copenhagen, Copenhagen, Denmark

John Stewart

Department of Physics and Astronomy, West Virginia University, Morgantown, West Virginia 26506, USA

Alexander Strahl

University of Salzburg, A-5020 Salzburg, Austria

Ana Susac

Department of Applied Physics, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, HR-10000 Zagreb, Croatia

Madison Swirtz

University of Utah, Salt Lake City, Utah 84112, USA

Joseph A. Taylor

Department of Leadership, Research, and Foundations, University of Colorado, Colorado Springs, Colorado 80918, USA

Adrienne L. Traxler

Department of Science Education, University of Copenhagen, Copenhagen, Denmark

Jingying Wang

Beijing Normal University, Beijing, China

Xiaomei Yan

Shanghai Jiao Tong University, Shanghai, China

Gábor Á. Zemplén

Institute of Business Economics, Eötvös Loránd University, Budapest, Hungary

Danhua Zhou

Beijing Normal University, Beijing, China

Dean Zollman

Department of Physics, Kansas State University, Manhattan, Kansas 66506, USA

Kristina Zuza

Department of Applied Physics, School of Engineering Gipuzkoa, University of the Basque, Country (UPV/EHU), Donostia, Spain

AIP PUBLISHING BOOKS

The International Handbook of Physics Education Research: Special Topics

TABLE OF CONTENTS

Introduction **xxix**

SECTION I: EQUITY AND INCLUSION IN PHYSICS EDUCATION RESEARCH

Section Editor: Geraldine Cochran

Chapter 1 **Teaching Physics with Disabled Learners: A Review of the Literature** **1-1**
Jacquelyn J. Chini and Erin M. Scanlon
1.1 Introduction 1-1
1.2 Methods 1-9
1.3 Findings 1-10
1.4 Discussion 1-25
1.5 Future Directions 1-29
References 1-31

Chapter 2 **Framework for and Review of Research on Assessing and Improving Equity and Inclusion in Undergraduate Physics Learning Environments** **2-1**
Sonja Cwik and Chandralekha Singh
2.1 Introduction and Framework 2-1
2.2 Quantitative Research on Physics Major-Level Equity in Outcome 2-7
2.3 Quantitative Research on Course-Level Equity in Outcome: Performance 2-8
2.4 Quantitative Research on Course-Level Equity in Outcome: Motivational Beliefs 2-11
2.5 Stereotype Threat 2-15
2.6 Qualitative Method Research 2-16

2.7 Interventions 2-18

2.8 Critiques, Ongoing Dialogues, and Future Directions 2-20

References 2-25

Chapter 3 **Research on Gender, Intersectionality, and LGBTQ+ Persons in Physics Education Research** **3-1**

Ramón S. Barthelemy, Adrienne L. Traxler, Jennifer Blue, and Madison Swirtz

3.1 Introduction 3-1

3.2 Avoiding Gap-Gazing 3-2

3.3 Intersectionality 3-5

3.4 LGBTQ+ Research in STEM Education 3-8

3.5 Conclusion 3-13

References 3-13

Chapter 4 **Research on Equity in Physics Graduate Education** **4-1**

Diana Sachmpazidi

4.1 Introduction 4-1

4.2 Research on Systemic and Cultural Factors as Barriers to Equitable Educational Experiences 4-2

4.3 Research on Student Experience in Graduate School 4-5

4.4 Implications for Researchers 4-8

Acknowledgments 4-9

References 4-9

Chapter 5 **Research Design Concerning Equity in Physics Education Research (PER)** **5-1**

Alexis V. Knaub and Lin Ding

5.1 Introduction 5-1

5.2 Making the Team: What to Consider When Initiating a Study on Equity 5-4

5.3 Developing Research Goals and Questions That Benefit the Population Being Studied 5-8
5.4 Recruiting and Selecting Research Participants, Particularly from Marginalized Populations 5-10
5.5 Research Instruments 5-13
5.6 Data Sources and Collection 5-15
5.7 Safety and Care for Researchers and Research Participants 5-18
5.8 Theoretical Frameworks, Data Analysis, and Interpretation 5-20
5.9 Sharing the Research Findings Beyond the Academy 5-22
5.10 Changes to Research Design 5-23
5.11 Discussion and Conclusions 5-24
References 5-25

SECTION II: THE HISTORY AND PHILOSOPHY OF PHYSICS IN PHYSICS EDUCATION

Section Editors: Donald J. Metz, Peter Heering, and Cibelle Celestino Silva

Chapter 6 Physics as a Human Endeavor 6-1
Richard Staley
6.1 Introduction 6-1
6.2 Early Physics Education Research: Diversity, Inclusion, and the Limits of the Physics Classroom 6-3
6.3 Combatting Bias: Gender 6-9
6.4 Combatting Bias: Race 6-13
6.5 The Cold War and the Foundations of the Modern Physics Discipline: Pedagogical Histories and the Anthropology of Physics 6-16
6.6 Climate Change and the Future Physics Classroom: Meeting the Anthropocene 6-20
References 6-23

Chapter 7 **The Role of Physics in Achieving Scientific Literacy in the Present and the Future** **7-1**
Hunkoog Jho
7.1 Introduction 7-1
7.2 Emerging New Literacies in the Societies 7-2
7.3 Scientific Literacy as a Key in the Present and the Future 7-5
7.4 Trends in Research on Scientific Literacy 7-7
7.5 Physics as a Key Content of Knowledge in Scientific Literacy 7-11
7.6 Physics as a Fundamental Belief in Scientific Literacy 7-14
7.7 Conclusion 7-19
Acknowledgments 7-20
References 7-20

Chapter 8 **History of Physics and Socio-Scientific Issues: Approaching Gender and Social Justice** **8-1**
Thaís Cyrino de Mello Forato, Isabelle Priscila Carneiro de Lima, and Gabriela Kaiana Ferreira
8.1 Introduction 8-1
8.2 Socio Scientific Issues in Science Education 8-5
8.3 SSI and HP to Promote Social Justice 8-11
8.4 Building an Example: Lise Meitner and SSI on Physics Teaching 8-16
8.5 Some Considerations 8-26
Acknowledgments 8-28
References 8-28

Chapter 9 **The Aims and Values of Physics** **9-1**
Andreia Guerra and Ivã Gurgel
9.1 Introduction 9-1
9.2 Historical Review of the Aims and Values of Physics 9-3
9.3 The Construction of Consensual Views of the Nature of Science 9-15

9.4 New Perspectives for the Role of History, Philosophy and Social Studies of Science in Physics Teaching 9-23

References 9-26

Chapter 10 Methods and Practices in Physics 10-1

Oliver Passon and Johannes Grebe-Ellis

10.1 Introduction 10-1

10.2 The Demarcation Problem and the Rise and Fall of the "Scientific Method" 10-3

10.3 Nature of Science and Scientific Inquiry 10-5

10.4 The History of Physics in Physics Teaching 10-8

10.5 Physics and Physics Education on Practices and Methods 10-12

10.6 Summary and Directions for Future Research 10-23

References 10-24

Chapter 11 Epistemic Beliefs and Physics Teacher Education 11-1

Gábor Á. Zemplén

11.1 Introduction 11-1

11.2 From the Emergence of Research on Epistemic Beliefs to Proliferation 11-2

11.3 Foundational Issues 11-16

11.4 On Some of the Many Roads Ahead 11-26

References 11-28

Chapter 12 Philosophy of Physics: Its Significance for Teaching and Learning 12-1

Roland M. Schulz and Calvin S. Kalman

12.1 Introduction 12-1

12.2 A Brief Review of Historical Interaction between Physics and Philosophy 12-3

12.3 Physics *and* Philosophy or Physics *without* Philosophy? 12-5

12.4 Philosophy of Physics (PoP) 12-9

12.5 Physics, Philosophy, and the Field of Education 12-13
12.6 Philosophy of Physics and PER: Learning Theories and Instructional Strategies 12-19
12.7 Conclusion 12-26
References 12-27

Chapter 13 New Methodological Approaches Toward Implementing HPS in Physics Education: The Landscape of Physics Education 13-1
Elizabeth Mary Cavicchi, Hillary Diane Andales, and Riley S. Moeykens

13.1 Introduction 13-1
13.2 Chapter Orienting Aspects 13-4
13.3 Widely Documented Trends 13-5
13.4 The Missing Student Perspective 13-7
13.5 Other Literature Reviews 13-8
13.6 Tours 13-9
13.7 Tour 1: Historical Experiments and Instruments 13-9
13.8 Tour 2: Contextualization through Narratives 13-14
13.9 Tour 3: Feminism and Indigenous Experience 13-20
13.10 Conclusion 13-27
Acknowledgments 13-27
References 13-28

SECTION III: PHYSICS TEXTBOOKS

Section Editors: Marika Kapanadze and Gabriela Jonas-Ahrend

Chapter 14 Expectations on Physics Textbooks 14-1
Sascha Grusche, Alexander Strahl, and Katrin Bölsterli Bardy

14.1 Introduction 14-1
14.2 Theory 14-2
14.3 Research Question 14-5
14.4 Methods 14-5

14.5 Results 14-9
14.6 Discussion 14-24
14.7 Conclusion and Outlook 14-32
References 14-33

Chapter 15 Textbook and Curriculum Alignment 15-1
Josip Slisko

15.1 Introduction 15-1
15.2 The Place and Role of Textbooks in an Extended Conception of Curriculum 15-3
15.3 Alignment of Physics Textbooks and Curricula in Physics Education Research: An Initial Contextual Consideration 15-5
15.4 Curriculum-Based Research of Physics Textbooks 15-9
15.5 Alignment of Physics Textbooks with Meta-Curricular Frameworks 15-19
15.6 Alignment of Physics Textbooks with Intra-Curricular Processes 15-26
15.7 Conclusions 15-30
References 15-32

Chapter 16 Analysis of Physics Textbook Content 16-1
Xiaomei Yan, Yuze He, Jingying Wang, Xiying Li, Xiaomei Ping, and Danhua Zhou

16.1 Introduction 16-1
16.2 Research Method 16-3
16.3 Basic Information about the Reviewed Literature 16-6
16.4 Findings on the Analysis of Physics Textbook Content 16-11
16.5 Content Analysis of Key Techniques and Methods of Physics Textbooks 16-24
16.6 Conclusions 16-27
16.7 Limitations and Implications 16-29
References 16-31

Chapter 17 **Evaluation of Physics Textbooks** **17-1**
Marika Kapanadze, Gabriela Jonas-Ahrend, Alexander Mazzolini, and Fadeel Joubran

17.1 Introduction 17-1
17.2 Methods 17-3
17.3 Findings 17-5
17.4 Conclusions 17-24
Acknowledgments 17-27
References 17-27

SECTION IV: MATHEMATICS IN TEACHING AND LEARNING PHYSICS

Section Editor: Gesche Pospiech

Chapter 18 **Role of Mathematics in Physics from Multiple Perspectives** **18-1**
Gesche Pospiech and Ricardo Avelar Sotomaior Karam

18.1 Introduction 18-1
18.2 Perspectives on the Interplay 18-2
18.3 Educational Perspective on the Relation of Mathematics and Physics 18-7
18.4 Frameworks for Describing the Blending of Physics and Mathematics 18-18
18.5 Conclusion and Outlook 18-24
References 18-25

Chapter 19 **The Meanings of Physics Equations in the Context of the Interplay between Physics and Mathematics** **19-1**
Minchul Kim, Yongwook Cheong, and Jinwoong Song

19.1 Introduction 19-1
19.2 Physics Equations in Physics Education 19-3
19.3 Perspectives on the Interplay between Physics and Mathematics 19-9

19.4 Students and Teachers Comprehension of Mathematization 19-13
19.5 Conclusion 19-19
References 19-20

Chapter 20 Graphs 20-1
Lana Ivanjek, Maja Planinic, and Ana Susac
20.1 Introduction 20-1
20.2 Theoretical Background 20-1
20.3 Graphs in Mathematics Education Research 20-4
20.4 Graphs in Physics Education Research 20-8
20.5 Instruments for Assessing Student Difficulties with Graphs 20-12
20.6 Interplay of Mathematics and Physics in Student Understanding of Graphs 20-16
20.7 Eye-Tracking Studies 20-19
20.8 Conclusion 20-21
References 20-22

Chapter 21 Visualization and Mathematization: How Digital Tools Provide Access to Formal Physics Ideas 21-1
Elias Euler, Lorena Solvang, Bor Gregorcic, and Jesper Haglund
21.1 Introduction 21-1
21.2 Theories of Visualization/Mathematization and Our Approach to Synthesis 21-3
21.3 Function I: Bridging Physical Phenomena and Formal Representations 21-8
21.4 Function II: Bridging Idealized Models and Formal Representations 21-13
21.5 Summarizing the History of Visualization Tools in Physics Education 21-20
21.6 Conclusions 21-21

Appendix—Legend for References in Fig. 21.8 21-21
References 21-23

SECTION V: PHYSICS EDUCATION RESEARCH: HISTORY, METHODOLOGIES, THEMES

Section Editor: David E. Meltzer

Chapter 22 **The Necessarily, Wonderfully Unsettled State of Methodology in PER: A Reflection** **22-1**
David Hammer

22.1 A Reminder of Methodology in Physics 22-3
22.2 Methodology in PER 22-4
22.3 Looking Forward 22-10
Acknowledgments 22-11
References 22-11

Chapter 23 **Toward More Rapid Accumulation of Knowledge about What Works in Physics Education: The Role of Replication, Reporting Practices, and Meta-Analysis** **23-1**
Joseph A. Taylor and Larry V. Hedges

23.1 Criticisms around Knowledge Accumulation for What Works in Education 23-1
23.2 Replication and Its Contribution to Knowledge Accumulation 23-6
23.3 Reporting Practices and Resources to Facilitate Replication 23-10
23.4 Institutionalizing Practices for More Rapid Knowledge Accumulation 23-16
References 23-18

Chapter 24 **Quantitative Methods in PER** **24-1**
John Stewart, John Hansen, and Lin Ding

24.1 Introduction 24-1

24.2 Research-Based Conceptual Instruments 24-2
24.3 Descriptive and Inferential Analyses 24-4
24.4 Classical Test Theory 24-9
24.5 Item Response Theory 24-12
24.6 Factor Analysis 24-15
24.7 Regression 24-18
24.8 Structural Equation Modeling 24-22
24.9 Cluster Analysis 24-25
24.10 Network Analysis 24-27
24.11 Machine Learning 24-28
24.12 Conclusion 24-29
Acknowledgments 24-29
References 24-29

Chapter 25 Qualitative Methods in Physics Education Research 25-1
Valerie K. Otero, Danielle Boyd Harlow, and David E. Meltzer

25.1 Introduction 25-1
25.2 Overview 25-2
25.3 Research Questions and Study Design 25-3
25.4 Interviews 25-6
25.5 Teaching Interviews 25-12
25.6 Surveys, Artifacts, and Electronic Sources 25-12
25.7 Observations 25-14
25.8 Processing Data 25-16
25.9 Coding and Analyzing Data 25-17
25.10 Theoretical Perspectives and Frameworks: Study Types 25-24
25.11 Validity and Reliability in Qualitative Research 25-29
25.12 Summary 25-31
References 25-31

Chapter 26 **Research-Based Teaching-Learning Sequences in Physics Education: A Rising Line of Research** **26-1**
Jenaro Guisasola, Kristina Zuza, Paulo Sarriugarte, and Jaume Ametller

26.1 Introduction 26-1
26.2 Research on the Design and Evaluation of Physics Teaching Learning Sequences 26-4
26.3 Design-Based Research: A Research Methodology on Teaching and Learning Specific Scientific Content 26-9
26.4 Defining Design Tools for Articulating Intermediate Frameworks 26-13
26.5 Exchange of Successful Educational Practices and Products 26-21
26.6 Perspectives of Design-Based Research and Evaluation of TLS 26-22
References 26-24

Chapter 27 **Epilogue** **27-1**
Dean A. Zollman

27.1 Introduction: A Little History 27-1
27.2 PER as Prologue to the 21st Century 27-3
27.3 Beyond the Skills 27-6
27.4 21st Century Graduate Education in PER 27-6
27.5 Conclusions 27-7
Acknowledgment 27-7
References 27-7

Index **I-1**

INTRODUCTION

The field of physics education research (PER) has been growing and diversifying over the past few decades. New conferences and new journals have emerged, and a brief survey of their contents reveals an intellectually and geographically diverse field of inquiry. More sophisticated methodologies are providing deeper insights into long-studied issues, and cross-pollination with other fields is opening up new avenues of inquiry. Critical reflections on the discipline of physics itself are causing researchers to question long-held traditions concerning not only how we teach, but also what—and who—we teach. In short, no aspect of the experience of students and their teachers is off limits, and we increasingly acknowledge the complex interplay between disciplinary culture, teaching environments and tools, and students' intellectual, social, and personal development.

While growth and diversification represent progress, they also present challenges. It is no longer possible for any individual to be aware of both seminal and pioneering work across PER, and to have a sufficiently deep grasp of methodologies to evaluate the trustworthiness of claims arising from distant areas. In some quarters, there is a feeling that the field is at a crossroads: we might consolidate around a few critical themes or evolve into a set of related sub-disciplines as has been the case with physics itself. Therefore, the moment seems right for scholars to reflect on the past, to synthesize what we have learned, and to look ahead to the future. *The International Handbook on Physics Education Research* serves as both a mechanism for such a reflection and a record of the results. The three volumes represent an up-to-date and authoritative review that encompasses all of the major strands of research. It is intended to help both newcomers and established researchers appreciate the major findings across all sub-domains, to discern global themes, and to recognize gaps in the literature. It is our hope that the *IHPER* will serve as a practical resource and contribute to vital conversations about what counts as PER, who counts as a physics education researcher, where we belong, and what, if anything, unifies us as a discipline.

In order to tackle this enormous challenge, we relied on the expertise—and enormous effort—of 19 leading scholars who served as section editors. They coordinated the construction of 12 sections covering Subject Matter Learning · Cognitive and Affective Aspects of Physics Learning Teaching · Educational Technologies · Physics Teaching Environments · Physics Teacher Education · Assessment · Equity · History and Philosophy · Textbooks · Mathematics · PER. The result is a coherent set of 69 chapters prepared, and reviewed, by a broad spectrum of established and emerging researchers representing different countries, career stages, identities, backgrounds, and perspectives. We know that the inclusion of diverse viewpoints has strengthened *IHPER*, ensuring it is valid, accessible, and relevant. We hope that the handbook will, in turn, contribute to the community's shared goals of greater equity and inclusion.

Contributors were asked to consider the following central questions:

- What has PER contributed to our current knowledge of teaching and learning of physics?
- What would we be lacking today without decades of continued PER?

- How has PER evolved over the decades (in terms of research questions, instruments employed, methodologies used, etc.)? What were the major turning points?
- How has physics teaching and learning changed over the decades due to the direct impact of PER?
- How has PER benefited from other disciplines (e.g. cognitive psychology, educational psychology, pedagogical research, instructional design research, etc.) and vice versa?

The resulting chapters present an authoritative overview of the literature with an emphasis on major achievements and highly influential studies. They also feature important work that is not widely familiar in the PER community. We hope that readers will be able to identify areas where the literature is sparse, conflicting, or out-of-date, or cases in which conclusions assumed to be well-established actually have thin support. Moreover, readers should be able to grasp not only where we are now, but how we got here.

The scope of the handbook is broad, but it does not attempt to survey all of physics education. In particular, we do not cover innovations in physics teaching that were not strongly driven by research or rigorously validated. Also, the handbook's focus is on *physics* education research. While science education research may be relevant, comprehensive handbooks already exist to survey that literature. For *IHPER*, the priority has been on research in which physics teaching is an intrinsic element. Finally, while the handbook is international in scope, the emphasis is on literature published in English, except for publications of exceptional significance.

By its nature, research on education cannot be neatly separated into clear categories. For example, research on the learning of specific content often overlaps with research on teaching strategies, which, in turn, often overlaps with research on assessment. Nevertheless, we have organized the handbook in three volumes. *Learning Physics* begins by focusing on the student and how they learn the specific concepts and practices of physics, including cognitive, affective and epistemological aspects of learning and teaching. *Teaching Physics* focuses more at the level of the classroom, with sections on technology, teaching environments, teacher preparation, and equity. *Physics Education Research Special Topics* features an even broader perspective including the impact of history and philosophy on physics teaching, and the role of mathematics in physics learning and teaching. *Physics Education Research Special Topics* ends with a more "meta" view of the field itself, including developments in what we study and how we study it. A final epilogue with reflections by Dean Zollman, a long-time leader in PER, concludes the handbook.

While we have aimed to be comprehensive, no handbook can do justice to an entire field, especially one as rich, diverse, and rapidly evolving as PER. Therefore, we anticipate that readers will identify gaps and omissions. We are aware of some already. The development of the handbook began in 2019, shortly before the global Covid-19 pandemic began. As months and years went by with no clear end in sight, some authors were forced to withdraw. A notice from the publisher of an unanticipated final deadline resulted in a flurry of activity to try to obtain reviews and finish chapters. Despite the valiant efforts of the section editors, authors, and reviewers, not every chapter could be completed in time. Nevertheless, we are confident that the PER community will find the handbook to be an invaluable resource, enriching research, provoking new strands of inquiry, and prompting reflection on our field.

SECTION I

EQUITY AND INCLUSION IN PHYSICS EDUCATION RESEARCH

Section Editor

Geraldine Cochran

Equity can be defined as a type of justice (Atwater, 2000). Historical and present inequities or injustice in education have been well documented (Ladson-Billings, 2006; Badat and Sayed, 2014; and Kranrattanasuit, 2023). Inequity in education exists throughout the globe. Thus, it is unsurprising—though disappointing—that inequity in physics education also exists throughout the globe (Hartline and Michelman-Ribeiro, 2005; and Cochran *et al.*, 2019). As a response to the lack of diversity and exclusioary practices utilized in physics, some have called for equity-oriented physics education research (Cochran *et al.*, 2020). Ironically, this section focuses on equity in physics education research and yet the composition of this section highlights inequity in physics education research. Whose work is deemed worthy of the classification of physics education research? Whose perspectives and experiences are included in physics education research? How do publishing practices (e.g., solicitations/recruitment, timelines, criteria etc.) create barriers for participation in the publication of physics education research? There are only five chapters included in this section. If inequity exists in physics education the world over, there should be many more chapters included in this section. The works cited should present a global perspective. I am grateful to the authors of the chapters included in this section as they used this opportunity to highlight inequity in physics education, explore how research has sought to address these inequities, and envisioned how physics education research might move forward in addressing existing inequities in physics education.

Inclusion is used to mean a variety of things in scholarly literature. In this section, we envision inclusion as centering people's experiences within the physics education environment (Cochran, 2018). The authors of the chapters in this section have not lost site of the importance of people's experiences. Indeed, they emphasize that the injustices identified have resulted in the minoritization and marginalization of people in physics education. Chapter 1 focuses on teaching physics with disabled learners, Chap. 2

focuses on the experiences of women and ethnic/racial minorities in physics in undergraduate physics learning environments, and Chap. 3 reviews research on gender, intersectionality, and LGBTQ+ persons in physics education. While Chap. 4 focuses on equity in graduate education, the authors have not lost sight of the experiences of those marginalized and minoritized in graduate education. Indeed, equity entails a redistribution of power and access to make situations and circumstances more just or fair (Atwater, 2000). Thus, this chapter highlights the inequitable access and experiences in graduate education in physics. The final chapter in this section, Chap. 5 examines equity considerations in the research design of all physics education research. Indeed, equity should be a consideration for all physics education research and not just research that focuses on specific injustices in physics education research or the experiences of marginalized people in physics/physics education.

REFERENCES

Atwater, M. M., Sci. Educ. **84**(2), 154–179 (2000).

Badat, S. and Sayed, Y., Ann. Am. Acad. Pol. Soc. Sci. **652**(1), 127–148 (2014).

Cochran, G., Scholarly Kitchen (published online 2018).

Cochran, G. L. *et al.*, *Handbook of Research on STEM Education* (Routledge, 2020), pp. 257–266.

Cochran, G. *et al.*, *Women in Physics: 6th IUPAP International Conference on Women in Physics, Birmingham, United Kingdom, 16--20 July 2017* (AIP Publishing, 2019).

Kranrattanasuit, N., Soc. Sci. J. **60**(1), 75–90 (2023).

Ladson-Billings, G., Educ. Res. **35**(7), 3–12 (2006).

Second IUPAP International Conference on Women in Physics, edited by B. K. Hartline and A. Michelman-Ribeiro (American Institute of Physics, 2005), Vol. 795.

CHAPTER 1

TEACHING PHYSICS WITH DISABLED LEARNERS: A REVIEW OF THE LITERATURE

Jacquelyn J. Chini[1] and Erin M. Scanlon[2]

Chini, J. J. and Scanlon, E. M., "Teaching physics with disabled learners: A review of the literature," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 1-1–1-34.

1.1 INTRODUCTION

The World Health Organization (WHO) defines disability as "an umbrella term, covering impairments, activity limitations, and participation restrictions" (World Health Organization, 2001). Furthermore, the WHO specifies that disability reflects "the interaction between features of a person's body and features of the society in which he or she lives. Overcoming difficulties faced by people with disabilities requires interventions to remove environmental and social barriers." This focus on the interaction of an individual with the physical and social environment is a hallmark of the *social model of disability* (Shakespeare, 2006). The social model of disability contrasts with the *medical model*, which emphasizes interventions aimed at changing or fixing the disabled individual, and other models that position disability and overcoming barriers as an individual's responsibility. To apply the social model to physics instruction, instructors should focus their attention, time, resources, and effort toward making the learning environments more accessible and inclusive, rather than trying to change individual disabled students.

Physics instructors should plan for learner variation when designing and implementing courses. However, research has shown that faculty across academic disciplines lack knowledge of accessibility laws (Thompson *et al.*, 1997; Zhang *et al.*, 2010; and Baker *et al.*, 2012), and do not feel prepared

---

[1] The first author identifies as a white cisgender woman with anxiety, depression, and obsessive-compulsive tendencies.

[2] The second author identifies as a white cisgender woman with migraines, anxiety, and depression.

https://doi.org/10.1063/9780735425514_001

The International Handbook of Physics Education Research: Special Topics, Mehmet Fatih Taşar and Paula R. L. Heron

to support disabled students[3] in their courses (Leyser *et al.*, 1998; Norman *et al.*, 1998; Reed *et al.*, 2003; Rao, 2004; and Evans *et al.*, 2017). Within STEM (science, technology, engineering, and mathematics) disciplines, research has found that STEM faculty are generally less amenable to the use of accommodations in their courses and hold more negative beliefs about disabled students than faculty in other academic disciplines (Schoen *et al.*, 1986; Lewis, 1998; Rao, 2004; and Skinner, 2007). Further, popular, research-based introductory physics curricula have not historically been designed to support disabled students or provide details for instructors about how to make modifications to support disabled students (Scanlon *et al.*, 2018). The purpose of this chapter is to review the literature related to teaching physics with disabled students to make suggestions for practice and for education researchers. The goal of this work is to move physics communities toward social justice and equity of access, support, and inclusion of all learners regardless of their disability status.

Postsecondary instructors typically do not know personal details about students, such as their disability diagnoses, which may seem like an impediment to reducing barriers in the learning environment. However, all people, regardless of their disability status, have a variety of needs, abilities, and interests (Scanlon and Chini, 2018). Design frameworks such as universal design for learning (UDL) broaden the instructor's focus from responding to emergent individual needs (e.g., via the use of university mandated accommodations) to proactive, inclusive design that inherently supports a wider variety of students without the need for specialized design or adaptation. To support this shift in framing, the findings in this chapter are sorted by instructional purpose and application rather than disability or impairment. The following subsections summarize the statistics about representation, legal requirements, experiences of disabled people in STEM, and state of access for disabled students in higher education.

1.1.1 Representation of disabled people in STEM

1.1.1.1 Population

Shifting definitions and ways of identifying disabled people complicate both representation estimates and identification of trends across time of the proportion of the population who identify with one or more disabilities. In 2011, the WHO estimated that 15% of people worldwide identify with a disability, an increase from prior reports that had the estimate at 10% (WHO, 2011). In 2016, the Census Bureau's American Community Survey estimated 11% of the U.S. working-age population identified with a

---

[3] Some people prefer person-first language (e.g., "students with disabilities," "person with visual impairment," or "scientists with a disability") because it emphasizes the person over the ability. However, others feel impairment-first language can highlight the social aspect of disability (e.g., inaccessible curricula create disabled students) and that the difference is an integrated part of the person's identity (e.g., Autistic person or Deaf person, just as we would typically say "tall person" rather than "person with tallness"). While there are trends in specific communities, there is not a single, universally accepted language related to disability. When you are talking with an individual, it is best practice to ask them about their preferred language. Language choices varied across the references reviewed; impairment-first language will be used in this chapter (except when directly quoting sources) and terms that have been identified as likely harmful were updated to reflect modern language.

disability, with the most common disability types described as "ambulatory" (5% of working-age population) and "cognitive" (4.5%; National Center for Science and Engineering Statistics, 2021). In the U.K. 18.6% of working-age women and 18.8% of working age men reported having a disability (Kirkup *et al.*, 2010). The Economic and Social Commission for Asian and the Pacific (ESCAP) states that the majority of ESCAP member States underreport the prevalence of disability in their population, with reports ranging from 1% in Laos to 24% in New Zealand (ESCAP, 2017). Researchers have struggled to find statistics in some regions, such as the European Union, Australia, and South Africa (Sukhai and Mohler, 2016). Yet, barriers to participation vary widely across cultures. For example, the World Report on Disability states that "many children drop out of school in Brazil because of a lack of reading glasses, widely available in most high-income countries" (2011). Disabled people represent a significant population worldwide.

1.1.1.2 Undergrad and graduate enrollment

The lack of statistics about disabled individuals across countries means little is known about disabled students' participation in undergraduate and graduate education globally. In 2016, 19.5% of undergraduate students in the U.S. reported one or more disabilities, and students who reported disabilities participated in undergraduate STEM education at a similar rate (28%) as students who did not report disabilities (National Center for Science and Engineering Statistics, 2021). Studies in the U.K. find a different trend at the postgraduate level, with "disabled STEM students 57% less likely to take up a postgraduate STEM study than non-disabled students" (Fell *et al.*, 1985).

As of 2019, Canadian researchers reported that statistics do not exist in the number of disabled students in postsecondary STEM majors or STEM occupations (Prema and Dhand, 2019). Following legislation mandating the provision of assistive devices in educational institutions, Slavin (2014) surveyed all postsecondary physics departments in Canada to benchmark the participation of legally blind students in postsecondary physics. No respondents were aware of low vision practicing physicists in Canada, and only two "legally blind" physics students were reported (Slavin, 2014).

In the United States, longitudinal studies such as the National Longitudinal Transition Study—2 (NLT-2) allow researchers to explore trends in undergraduate student populations. Using the NLT-2, Lee (2011) found that disabled students were more likely to enroll at two-year colleges than non-disabled students, and at two-year colleges, disabled students were more likely to enroll in STEM majors than non-disabled students. Additionally, disabled students in STEM majors reported receiving fewer accommodations than disabled students in non-STEM majors across all institution types in the U.S. This discrepancy in accommodation use could be due to the attitudes and beliefs of STEM faculty (who have been shown to be less willing to provide accommodations than their colleagues in other academic disciplines), the nature of the course (e.g., physics instructors could experience difficulties knowing how to accommodate in a lab setting or with mathematical representations), and/or students requesting less accommodations. In further analysis of the NLT-2 data set, Lee (2022) found that students enrolled at a two-year college who identified as having a "a problem conversing" were nearly

5 times more likely to enroll in a STEM major than those who did not identify as having difficulty conversing (Lee, 2022). Additionally, disabled students from lower economic backgrounds were more likely to enroll in a STEM major than disabled students from higher economic backgrounds. The National Science Foundation (NSF) in the U.S. reports that students who reported having one or more disabilities were likely to be older than their peers who did not report disabilities (National Center for Science and Engineering Statistics, 2021). Overall, disabled students represent a sizable fraction of postsecondary STEM students.

1.1.1.3 Employment and salary

In the United States, approximately 10% of employed scientists and engineers identified with one or more disabilities, with reported disability rates somewhat higher among men than women and somewhat lower among Asian scientists than non-Asian scientists; both trends are possibly related to the relative age of these demographic groups (National Center for Science and Engineering Statistics, 2021). However, workforce studies have found that while a much higher percentage (34%) of employees would fit current U.S. federal definitions of disabled, only one-third would disclose their disability status to their employer (more in line with the 10% of employed scientists and engineers cited earlier) and fewer would disclose to their colleagues (Sherbin *et al.*, 2017; and Jain-Link and Kennedy 2019). Disabled individuals who intend to join the STEM workforce are more likely to be employed (65%) than disabled individuals in the U.S. overall (32%; Lee, 2022). In the U.K., disabled individuals comprised about 10% of the science, engineering, and technology workforce in 2003 and 2008 (Kirkup *et al.*, 2010). Median salaries are about the same for disabled and non-disabled individuals within science and engineering occupations, and median salaries for science and engineering occupations are significantly higher than those for non-science-and-engineering occupations (National Center for Science and Engineering Statistics, 2021). STEM careers can provide a path to economic security. Since STEM education is the main mechanism by which individuals join the STEM work force, it is essential that STEM education supports disabled students.

1.1.2 Legal requirements

While legal requirements vary by country (Sherbin *et al.*, 2017), several international standards shape local laws. The *United Nations Convention on the Rights of Persons with Disabilities* (2008) states that co-signing countries should “ensure an inclusive education system at all levels and lifelong learning” to support the development of a sense of dignity, self-worth, personality, talents, creativity, and mental and physical abilities, such that disabled people can “participate effectively in a free society.” The United Nations catalogs national disability laws and acts (United Nations, n.d.). The Web Content Accessibility Guidelines (WCAG), developed by the World Wide Web Consortium (W3C), provide a universally accepted set of digital accessibility guidelines which have been used as accessibility standards in national laws (see WCAG Web Accessibility Laws & Policies site, for examples; Worldwide Web Consortium, 2019).

However, as Prema and Dhand (2019) explicate in a Canadian context, the existence of laws does not directly translate to full inclusion of disabled students in STEM education. In Canada, disabled individuals are protected from discrimination by "quasi-constitutional" human rights legislation, such as the provincial *Human Rights Code* and *the Canadian Charter of Rights and Freedoms*, which includes equal access to education. These common laws and codes require "post-secondary institutions to accommodate students with disabilities until undue hardship," which courts and tribunals have assessed via "cost, external sources of funding if any, health and safety requirements" (Prema and Dhand, 2019, p. 129). However, "educators are often unsure of how to apply the legal requirements of the duty to accommodate appropriately for students pursuing STEM, while balancing the factors of health, safety and cost" (p.123).

Prema and Dhand explain that "Canadian human rights codes fail to create 'positive obligations' which ensure inclusion and accessibility within post-secondary institutions" (Flaherty and Roussy, 2014, p. 8). Instead, the legislative framework sets up complaint procedures, mechanisms for accommodations if requested, and compensation for past wrongs in cases of discrimination (Flaherty and Roussy, 2014). As Flaherty and Roussy (2014) suggest, this leads to an '*ad hoc* enforcement of human rights,' which is described as the following: '[T]he onus of asserting rights or identifying Code breaches rests with students. In a manner of speaking, this leads to an *ad hoc* enforcement of human rights, where only those who complain see their rights enforced. As a result, those students who lack the will, endurance, means or ability to lodge a formal complaint may continue to be victims of discrimination' (p. 8)" (Prema and Dhand, 2019). Thus, physics instructors must go beyond the minimal legal requirements to support disabled students.

1.1.3 Overall trends in STEM culture toward disabled people

In the last half century, there have been myriad reports and guides designed to assist instructors and administrators in supporting disabled students in the postsecondary physics setting. This section includes a belief overview of these materials. In the early 1980s, an NSF-funded project employed a critical incident technique to collect examples of (in)effective instruction experienced by blind students in postsecondary STEM in southern California for the purpose of improving STEM instruction for visually impaired students. Interviews with 105 blind students revealed effective teaching practices that provided access to information, enhanced motivation and interest, and allowed for social interaction and flexibility with time. These teaching practices include concrete learning experiences (i.e., relating concrete models and materials to abstract concepts), creative use of learning materials (i.e., field trips, multisensory learning experiences), and detailed descriptions and explanations (i.e., teacher clearly verbalizes visual information, such as writing and images). Ineffective teaching behaviors were described as the absence of these effective behaviors (Sica, 1982).

In the early 1990s, Sheryl Burgstahler, writing in the context of the NSF-funded University of Washington DO-IT (disabilities, opportunities, internetworking, and technology) program, described the main factors leading to underrepresentation of disabled people in science, engineering, and math as "preparation of students with disabilities; access to facilities, programs, and equipment; and acceptance by educators, employers and co-workers." Burgstahler proposed solutions such as encouraging disabled students to be self-advocates, encouraging them to take high school science and math courses and connecting students with disabled role models. Burgstahler (1994) called for increased access to technology for disabled students. However, she still identified negative attitudes as "the single most significant barrier faced by individuals with disabilities pursuing careers in science and engineering."

In the late 1990s, with support from NSF and American Association for the Advancement of Science (AAAS), Seymour and Hunter (1997) conducted a study to contribute to the ongoing debate about the cause of underrepresentation of disabled students in STEM. Through interviews and focus groups with a total of 65 disabled students, the authors investigated their education and work experiences. They found that "given the many types and degrees of medical conditions which are encompassed by the term 'disability,' one way to understand the commonality of their experience is to see all students with disabilities as students who are 'time-disadvantaged'" (p. 167), meaning that their impairment(s) coupled with their learning experiences cost them more time than their non-disabled peers. Faculty and STEM professionals held narrow ideas of the time required to engage in and complete tasks that were not inclusive of the needs and abilities of the disabled participants.

In 2014, the AAAS published "Fostering Inclusion of Persons with Disabilities in STEM," which focused on four broad topics: (1) facilitating disabled student participation with technology; (2) interventions for college students to enhance retention, persistence, and career readiness; (3) dissemination of evidence-based technologies and methods for supporting disabled students; and (4) sustainability of programs for disabled students. Thus, the conversation about how to support disabled students in STEM has not changed much since the 1990s.

In 2008, the Institute of Physics (IoP), a professional society based in the U.K., published *Access for all: A Guide to Disability Good Practice for University Physics Departments* (Institute of Physics, 2008). *Access for All* describes the main barriers to participation in physics for disabled individuals as environmental, institutional (i.e., admissions policies and teaching methods), and attitudinal. The guide points out that in the U.K., the Disability Equality Duty "requires universities to be proactive in ensuring that disabled people are treated fairly" (p. 7). "This means that universities and departments must anticipate the general requirements of disabled people with a range of impairments and health conditions rather than waiting until a disabled person requests a particular adjustment. There is no defence for not making a 'reasonable adjustment'. If an adjustment is 'reasonable,' then it must be made" (p. 12). Additionally, the guide states, "To ensure that they are not discriminating against disabled applicants, universities must be able to demonstrate that the competence standards that they use for selection are appropriate and necessary; applied equally to disabled and non-disabled applicants; a proportionate means to achieving a legitimate aim" (p. 19). Thus, the guide offers a cultural shift toward anticipating the range of needs that contrasts sharply with the "*ad hoc* enforcement of human rights" described by Flaherty and Roussy (2014) a few years later in Canada.

In 2016, Sukahi and Mohler, both disabled scientists themselves, wrote *Creating a Culture of Accessibility in the Sciences*, which is a book targeted at higher education faculty, administrators, and employers across the STEM fields. This book provides insights and discussions of best practices to increase the accessibility of science for disabled people. *Creating a Culture of Accessibility in the Sciences* approach is novel in that it provides a comprehensive list of practices coupled with a suggested roadmap that higher education and industry professionals can implement to support disabled people. Additionally, Sukhai and Mohler describe the multiple roles that disabled students often take in the sciences: student as an educator (focusing on the role of students in self-advocating and educating others about disability topics), student as learner (focusing on how to support disabled students in postsecondary science courses), and student as mentee, trainee, and leader (focusing on how to support disabled students in research settings).

In 2021, the American Association of Physics Teachers (AAPT) published a white paper commissioned by its Committee on Laboratories called *Increase Investment in Accessible Physics Labs: A Call to Action for the Physics Education Community* (Dounas-Frazer *et al.*, 2022). The authors "call on the physics community to invest time, energy, and resources to increase the accessibility of undergraduate physics labs" (p. ii) and include a list of ideas for investment, testimonies from current and former disabled physical science students, a glossary of common disability terms, and appendices written by disabled students with suggestions of how to support students' specific impairments. Thus, the call-to-action cast students in all three roles identified by Sukahi and Mohler, learners and trainees in undergraduate labs as well as educators who can help the physics community better support disabled students.

Overall, recent reports have trended toward proactive support for disabled people in STEM; yet, much work is still needed. The STEM community should continue to learn from the disability community and shift from "ad-hoc enforcement of human rights" to a "positive obligation" to fully include disabled individuals.

1.1.4 Prior literature reviews and summaries of disability in STEM education

In 1994, the Science Association for Persons with Disabilities (1994) published a bibliography of over 1000 publications related to teaching science to disabled students. In the nearly thirty years since this publication, language around disability has changed, yet many of the topics highlighted in the bibliography remain prevalent in the literature today, such as teaching students with specific impairments (e.g., blind students, deaf and/or hard of hearing students, autistic students, and cognitively impaired students), technology-assisted instruction, and inclusion of disabled students in all educational settings.

In 2010, Leddy wrote an overview of the National Science Foundation's Research in Disabilities Education (RDE) program highlighting the need for rigorous research designs to examine the efficacy of technologies to support learning, degree completion rates, and transition to the STEM workforce for disabled individuals. An external evaluation across the program identified practices that contribute to the persistence of disabled students in STEM degree programs, including financial support, cooperative learning experiences, research experiences, off-campus externships, mentoring, and participation in STEM clubs, activities, and learning communities. However, open questions remained about the format of mentoring and optimal match between the mentor and mentee. While many projects have developed accessible technology to engage learners in STEM, the technologies had not been broadly adopted by high school and postsecondary education (Leddy, 2010). Thurston *et al.* (2017) further described findings from their synthesis of the NSF-RDE program. They identified common challenges including that disabled students did not receive adequate preparation for postsecondary STEM courses due to low expectations and insufficient access; lack of understanding, knowledge, skills, and cooperation from administrators, faculty, and staff; lack of accommodations such as accessible technology, accessible buildings, and accessible learning spaces; and lack of identifying, recruiting, and tracking disabled students to measure program impacts. Thurston *et al.* highlight successful practices, including engaging campus disability service offices to provide accommodations; cataloging and using existing campus and community resources before developing new resources; using a variety of recruitment and support strategies; and providing professional development for faculty staff, and administrators, and support for universal design for learning. PIs of RDE projects also had suggestions for facilitating a cultural shift in faculty and staff toward more positive attitudes and beliefs about disabled students, including adopting the socio-cultural model of disability and “using ‘PR’ campaigns about the strengths of students with disabilities” (p. 56).

In line with Leddy’s (2010) recommendation for rigorous research designs, Schreffler *et al.* (2019) conducted a systematic review of empirical literature published in peer-reviewed journals between 2006 and 2019 on UDL in postsecondary STEM. They identified four studies and three literature reviews. Thus, while some researchers have begun to use rigorous methods and examine UDL rather than solely accommodations, there is still a dearth of empirical literature to support best practice. Another systematic literature review, conducted by Kolne and Lindsay (2020), focused on peer-reviewed articles published between 1993 and 2008 reporting an empirical investigation of STEM interventions for disabled students. Kolne and Lindsay identified a small number of publications ($N = 17$) that met their inclusion criteria. Kolne and Lindsay state that the strongest evidence was found in two studies of virtual mentoring programs “in the context of perceived self-efficacy and in combination with STEM-specific training” (p. 541). Positive outcomes were reported for STEM interest, pursuit of STEM education and careers, and participants’ self-concept, as well as for all course-based interventions. Kolne and Lindsay reiterate the call for more rigorous, controlled research designs and examination of specific intervention components as well as raise the need for studies to explore issues of intersectionality (i.e., disability and gender) and the effects of interventions in various educational settings and countries.

Applying a different framing to assess the state of STEM education research, Li *et al.* (2020) analyzed trends in STEM education projects funded by the U.S. Institute of Education Sciences (IES) from 2003 to 2019 ($N = 127$). IES specifically funds research in special education, a term from K-12 education that originally referred only to disabilities but has since broadened to include other populations, such as English-language learners and students from low socioeconomic backgrounds. The researchers found that 28 projects in the “Special Education Research” category focused on disabled individuals and identified three relevant projects in the “Education Research” category, accounting for 24.4% of the projects. Across both funding programs, Li *et al.* found that the majority of funded projects were “development and innovation” (i.e., focused on developing new interventions; 45.7%), followed by “efficacy and replication” (i.e., focused on investigating impact; 26.8%) and “measurement” (i.e., developing and revising assessments, 16.5%). Longitudinal trends suggest a possible shift toward efficacy and replication studies, perhaps in response to calls for such work, as described above. Li *et al.* concluded that “Research on STEM education with special participant populations is important and much needed. However, related scholarship is still in an early development stage” (p. 9).

Traxler and Blue (2020) synthesized disability frameworks to “distill themes to guide the study of disability in physics,” which they identify as essential since such frameworks are “deeply embedded and implicit” in doing physics education research (PER) (p. 132). Frameworks differ in where disability is placed (i.e., an individual condition solved via individual intervention in the medical model, vs an interaction between a person’s impairments and social structures, solved via social design, in the social model). Traxler and Blue elevate the importance of being precise about the goals of research, contrasting accessibility research guided by the hope that someday no accommodations will be needed while valuing “neurodiversity and the diversity of bodies” (p. 139). The DisCrit (Annamma *et al.*, 2013) framework amplifies the importance of intersectionality, or the effect of combinations of identities, as “diagnoses and experiences of disability play out in racialized ways” (Traxler and Blue, 2020; p. 143). Traxler and Blue summarize several key ideas that should shape the future of research on disability in physics: disability is interlinked with other facets of identity, question “who gets to belong” and “who is normal,” and the importance of telling one’s own story about one’s self. Traxler and Blue’s discussion provides key ideas that can guide PER in the future.

Overall, this review of the state of STEM education research on disability indicates that not much has changed in the last fifty years. There is a need for both empirical research that evaluates the efficacy of interventions across learner populations and educational settings as well as explicit use of frameworks to shape research and detail beliefs and assumptions about disability. The purpose of this chapter is to present and critique extant literature at the nexus of physics, teaching, and disability.

1.2 METHODS

1.2.1 Selection of the articles

Several methods were used to identify sources to include in the chapter. Physics education journals (i.e., *American Journal of Physics, The Physics Teacher, Physics Education, Physical Review Physics*

Education Research) were Boolean searched with the keyword "disab*." Then, Google Scholar was used to search for "physics OR STEM AND disab*." To broaden the search, the references cited by the identified publications were examined, and Google Scholar was used to identify publications that had cited the identified publications. Overall, we searched for sources from November 2020 through June 2021 and identified 205 potential sources. The term "sources" is used as sources beyond journal articles, such as reports and dissertations.

Next, a portion of the sources were reviewed to define exclusion criteria. For this chapter, the exclusion criteria were

- sources focused only on K–12 education with no significant discussion of higher education or the STEM workforce (28 articles removed);
- sources focused only on another STEM discipline, including astronomy or pre-service teachers (if source focused on broad science, it must specifically focus on physics as well; 45 articles removed);
- sources which did not have sections focused on disability or have significant findings or discussion about disability (9 articles removed).
- sources which did not focus on teaching and learning (e.g., campus-wide support programs; technology/equipment without examples for physics teaching; 14 articles removed);
- sources other than articles, reports, dissertations, and book chapters (e.g., personal websites). Additionally, sources that the authors could not locate in English were not included (11 articles removed)
- sources that did not have implications for instructional practices and/or physics education research (34 articles were removed, many of which were included in the introduction and/or future direction sections).

After applying the exclusion criteria, 66 sources remained. These sources were then reviewed and sorted by audience (conducted by/for education practitioners or by/for STEM education researchers) and topic (laboratory practice, general education practice, technology, conceptual understanding, and universal design for learning). Identification of related literature was challenging because few articles cited related extant sources (i.e., the extant literature is not a well-connected network).

1.3 FINDINGS

This chapter summarizes 66 sources that describe education and research at the nexus of physics, teaching, and disability. The earliest of these sources was published in 1965 and the latest in 2021, with a median publication year of 2014 and an average of 2008. The sources focused on a variety of disabilities/impairments. Sixteen sources focused on disability in general without disaggregating impairment types, and 8 focused on multiple categories of impairment. Many sources focused on specific impairments, including 29 focused on visual impairments (i.e., blind, low-vision, screen-reader user), 6 focused on cognitive impairments [i.e., attention-deficit hyperactivity disorder (ADHD), learning disabilities, autism spectrum disorder, intellectual disabilities, developmental disabilities], 4 focused on hearing impairments (i.e., deaf, hard-of-hearing), 6 focused on multiple types of impairments, and 3 focused on physical/mobility impairments (i.e., mobility impaired, wheelchair users). None of the sources focused on health or emotional/mental health impairments.

Additionally, the sources were of many different types, including journal articles (49), books and book chapters (4), reports (2), conference proceedings (10), and dissertations (1). Journal articles have been published in many journals, including *The Physics Teacher* (6 articles), *Physics Education* (6), and *Proceedings of the Physics Education Research Conference* (5). Five additional journals each published two articles, and an additional 20 journals each published one article. The sections below include a summary of the findings and suggestions for practice and research from these 66 sources. The findings are disaggregated by the audience. The findings for practitioners are written to provide concrete suggestions for practice, whereas the findings for researchers include methodological information and suggestions for researchers.

1.3.1 Findings for practitioners

This section includes articles written for and by practitioners. The main emphases of these articles are instruction and include suggestions for practice. Many of the identified sources described teaching strategies that the authors had used and/or developed to include disabled students in their physics courses. The subsections below present a review of this literature disaggregated by the aspect of the course the source focused on, including the laboratory setting, lecture demonstrations of physics content, virtual simulations of experiments and concepts, lecture and direct instruction strategies, textbooks, and general inclusive instructional practices. This section is not disaggregated by disability or impairment because there are suggestions for practice that span disability types.

1.3.1.1 Laboratory setting

Descriptions of how to modify existing laboratory equipment to include disabled students in the physics lab setting was the most commonly discussed course aspect (24 of the 66 sources). Sources discussed multiple ways to ensure the laboratory environment and experiments are accessible to students who were categorized as (a) modifications to existing equipment; (b) accessible laboratory tools and assistive technologies; and (c) methods and tools to make specific experiments accessible.

1.3.1.1.1 Modifying existing equipment

In the first category of modifications to existing equipment, sources written over five decades describe how to make physics laboratory equipment accessible to visually impaired students (Henderson, 1965; Baughman and Zollman, 1977; Weems, 1977; Gough, 1978; Stewart, 1980; Cetera, 1983; Windelborn, 1999; and Brazier *et al.*, 2000). Suggestions synthesized across these sources for supporting visually impaired students include

- using raised lines to provide access to diagrams and graphs (Henderson, 1965; Weems, 1977; Stewart, 1980; Cetera, 1983; Windelborn, 1999; and Brazier *et al.*, 2000);
- creating tactile representations of key features of laboratory equipment (e.g., tactile metersticks or micrometers which allow users to feel measurement readouts; Baughman and Zollman, 1977; Weems, 1977; Cetera, 1983; and Windelborn, 1999);
- using magnification of measurement readouts (e.g., large print; Henderson, 1965);
- using audification (converting output to sound) of measurement readouts (e.g., talking clocks or calculators; Henderson, 1965; Weems, 1977; Cetera, 1983; and Windelborn, 1999);
- adding Braille writing to equipment instructions and readouts as well as demonstrations (Henderson, 1965; Baughman and Zollman, 1977; Weems, 1977; Gough, 1978, Stewart, 1980; and Cetera, 1983).

Relatedly, Supalo *et al.* (2007) described the programming modifications required to make Vernier's Logger Pro compatible with a common screen-reader software called JAWS (Job Access With Speech). Screen-reader compatibility is crucial to support visually impaired students. Supalo *et al.* provide a key connection between data acquisition tools that are commonly used in introductory physics laboratory courses and commonly used assistive technologies. Thompson (2005) focuses on providing access to LaTeX files for visually impaired students through the use of LaTeX2Tri. This tool allows users to input TeX, Word, and PDF files and the tool converts them to WinTriangle, which the author states is "the working language of many blind or visually impaired students and researchers…completing the loop of mathematical communication between the blind and sighted communities" (p. 1).

Along the same vein, Azevedo and Santos (2014a, 2004b) describe modifications to optics equipment that can be made to support visually impaired students. Specifically, the authors describe how ray tracing diagrams can be created via magnets representing the ray, optical axis, object, and image, and a magnetic board to support students to tactilely engage with the diagrams. Similarly, de Azevedo *et al.* (2015) suggest shining laser beams on students' hands or arms in order to allow them to feel the laser beam as a means to provide tactile access to laser light. The authors also include safety information for shining laser beams on the skin.

To support students with physical/mobility impairments (i.e., wheelchair users in this study), Bernhard and Bernhard (1998) discuss the feasibility and advantages of using microcomputer-based labs (MBL) where the digital data collection is possible. Nowadays, the use of computers and digital data acquisition tools (e.g., Pasco and Venier products) is common, and with small modifications could be used to support disabled students. Similarly, Frinks (1983) identified two accommodations required to support a wheelchair-user in accessing introductory physics labs: the table heights and utility access controls should be altered so that the student could access and engage with the tools while seated in their wheelchair. These accommodations nowadays are commonly incorporated into building designs due to the prevalence of universal design in architecture as well as local legal requirements (e.g., the Americans with Disabilities Act in the U.S.).

1.3.1.1.2 Accessible laboratory tools

In the second category, sources describe specially designed accessible laboratory tools to provide access to the experiment, equipment, and/or data for disabled students. Carver (1967) describes the design and implementation of a light probe that can be used by visually impaired students to detect motion. In particular, the author described the design and circuitry of such a light probe, as well as a short description of how to use the light probe. The light sensor can be "focused at short object distances as a microscope, at intermediate distances as a 'flag' for moving objectives, and at infinity for certain optical experiments" (p. 61).

Van Domelen (1999) introduced an artificial right-hand rule device that is made of a clear, plastic rectangular prism with vector arrows on three sides. This tool can assist students who have dexterity difficulties, students without right hands and/or the fingers used in the rule, and other students who have difficulty visualizing the three-dimensional vectors involved in the rule.

Tomac *et al.* (2016) use wooden blocks that have been calibrated to correspond to smaller distances to replace calipers to support visually impaired students. To provide virtual access to in-person experiments using microscopes, Mansoor *et al.* (2009) created the AccessScope application that allows students to remotely access and operate a microscopy workstation. This is especially important in the COVID/post-COVID era where people do not travel as frequently and others at high risk of illness limit their exposure. To support students with low vision, Cole and Slavin (2013) describe a video assistive device that allows users to view laboratory equipment or text by magnifying an image of the target. To help blind and sighted students learn the differences between displacement and distance, Bülbül *et al.* (2013) created a tool using a CD a string. The string is pulled across the CD and is used to measure the displacement of an object moving around the perimeter of the CD. This tactile tool is accessible to sighted and visually impaired students.

In the same vein, there are articles that focus on how to make a piece of equipment accessible for specific groups of disabled students. For example, Negrete *et al.* (2020) describe the use of a dial with slits and a photogate sensor to allow data audification for visually impaired students. Specifically, the rotating dial periodically blocks the photogate sensors. The photogate sensors are then connected to a device that converts the photogate signal to sound.

1.3.1.1.3 Methods and tools for specific experiments

In the final category, additional sources discuss how to make a specific experiment and/or topic accessible to students. To support wheelchair-users, Bernhard and Bernhard (1999) describe an experiment where students use wheelchairs on ramps with a motion detector to help students understand basic kinematics. Specifically, students measure the motion of a wheelchair-user rolling down a ramp and discuss the kinematics and/or forces. Similarly, Bülbül (2009) discusses access for visually impaired students when learning about optics. The author developed instructional materials called KAGOAD (Küresel Aynalarda Görüntünün Oluşumunu Anlatan Düzenek, which translates to the mechanism describing the formation of the image in spherical mirrors in English), which uses tactile representations of curved mirrors and light rays. These representations involve a foam board, needles, string to represent the light rays, and sugar cubes to represent the object and image.

To support hearing impaired students, Truncale and Graham (2014) described an experimental setup aimed at allowing hearing impaired students to engage in a sound laboratory focused on determining and plotting hearing sensitivities. In this article, Truncale and Graham describe an electro-optical eardrum that measures vibrations of a synthetic eardrum membrane via a laser. This allows students to engage in the activity without requiring the use of hearing.

1.3.1.2 Virtual simulations

There are many reasonable and appropriate reasons why an instructor would want to allow students to engage with a virtual simulation of an experiment or concept instead of a hands-on laboratory including (a) accommodating a student who has to miss class and/or lab; (b) teaching via remote instruction (as was commonly required due to the Covid-19 pandemic); (c) to support students with physical, dexterity, and/or mobility impairments whose access to the laboratory equipment is not supported; and (d) for students with attention difficulties (such as ADHD) to allow them to rework through the laboratory at their own pace.

There are numerous articles in the extant literature about how to support students via one platform of virtual simulations called Physics Education Technology (PhET) simulations. The PhET research and development team recently launched an accessibility initiative with a goal to provide access to the simulations to a wide range of users. The PhET team has written about the following accessibility features.

- screen-reader compatibility (Smith *et al.*, 2016a, 2016b; and Smith *et al.*, 2017);
- alternative keyboard compatibility and navigation access (Moore and Perkins, 2018);
- auditory descriptions of simulation design scenarios to support a wide range of users (Moore *et al.*, 2018; Moore and Perkins, 2018; Tomlinson *et al.*, 2018, 2019; and Winters *et al.*, 2018);
- data sonification (Moore and Perkins, 2018; and Tomlinson *et al.*, 2019).

As of the writing of this article, there are 33 PhET simulations about a variety of physics and chemistry topics that include at least one accessibility feature. There is one additional article written about non-PhET simulations. Farrell *et al.* (2001) focused on accessibility features of a spring force simulation that utilized force feedback (i.e., feedback given to the user about the strength of a force via motions of the mouse). In this study, the authors simulated the relationships between a spring length, applied force, and spring constant. The authors describe the effect of the force feedback on visually impaired and non-disabled students.

1.3.1.3 Direct instruction

There are few sources that describe best practices for direct instruction (i.e., lecture, didactic instruction) in physics courses and all of these sources center on visually impaired or hearing impaired students. Across these sources, the literature suggests

- using tactile, three-dimensional representations during class to help describe physics concepts to support visually impaired students (Sevilla *et al.*, 1991).
- allowing visually impaired students to use the accessibility tools of their choice. Parry *et al.* (1997) found that one student preferred to use Braille to conduct calculations, while another preferred to do calculations in their heads. Lannan *et al.* (2021) also suggest that instructors provide support and training for students about how to use these assistive tools.
- converting all visual course material into an accessible format, such as audible or tactile formats. Holt *et al.* (2019) also suggest that instructors should work with the visually impaired student and a staff person from the local disability services office to identify and address the individual needs of students.
- using visual and tactile modalities in place of auditory information to support hearing impaired students (Lang, 1973).
- using "See and Feel" sensory-focused pedagogy (i.e., where students will see and feel a phenomona) to provide access to lecture material and demonstrations for hearing impaired students (Vongsawad *et al.*, 2016).

1.3.1.4 Demonstrations

Creating accessible demonstrations to be showcased in the lecture, recitation, and/or laboratory setting is also important to ensure equitable access to the course. Three sources were identified discussing demonstrations of physics concepts, and they all focused on supporting access for visually impaired or hearing impaired students. Goncalves *et al.* (2017) pose a demonstration of the relationship between period and length for a pendulum by converting the pendulum bob's position to a sound frequency (via the use of an Arduino, an ultrasonic sensor, and a speaker).

Lang (1981) and Vongsawad *et al.* (2016) both describe how to provide access to sound concepts for hearing impaired students. Lang suggests the use of an oscilloscope to provide access for hearing impaired students to a Kundt tube and suggests using a ripple tank to showcase the Doppler effect. Vongsawad *et al.* describe the use of a Ruben tube demonstration (composed of a speaker changing the pressure of gas in a metal pipe with holes at the top, creating dancing fire standing waves) and suggest the use of Chladni plates connected to an accelerometer whose output (and the frequency) displays on a screen can provide access to vibrational modes to hearing and visually impaired students.

1.3.1.5 Textbooks

Many instructors supplement the information provided to students in class via the use of textbooks. While there are numerous textbooks available to cover introductory physics content, not all textbooks have the same level of accessibility and/or support for disabled students. For example, many physics textbooks use diagrams and graphs to represent information and relationships. However, this visual information is not natively accessible to visually impaired students.

Dickman *et al.* (2014) describe how to adapt physics diagrams for visually impaired students by converting the information from visual to tactile representations. The authors created tactile symbols for common elements in mechanics topics such as vectors, ropes, blocks, and pulleys. They then conducted a study of the effect of the tactile representations for three blind students and found that after sufficient training, the students were readily able to identify the representations and often did not require a spoken description of the diagram in order to understand what was represented. Similarly, Torres and Mendes (2017) describe a similar method for converting visual diagrams to tactile representations. However, they do not create tactile representations for an entire element (e.g., a pulley) but instead use KitFits which include general shapes (e.g., circles, rectangles, triangles, lines, curves) that can be used to create elements such as pulleys or vectors.

Kouroupetroglou and Kacorri (2010) describe an extensive process that can be used to convert inaccessible electronic copies of textbooks into multiple more accessible formats (i.e., Braille, audio-tactile, digital audio, and large print). The authors focus on creating accessible versions of mathematical and scientific expressions as well as visual diagrams, graphs, and graphics. The authors also suggest the use of universal design for learning (described in more detail in Sec. 1.3.1.6) as a framework for how to work toward making course materials more accessible and as a means to provide options and support for students.

1.3.1.6 General instructional practices

Lannan *et al.* (2021) discuss general accessibility tools that can be used in the laboratory and lecture setting for a wide range of students. In particular, the authors suggest instructors consider the universal design for learning framework and state: "the first step to implementing universal design is to examine the why, what, and how of our teaching while looking for the barriers our students frequently encounter. Instructors should ask themselves: 'Why should students care about this topic?'; 'What do students find challenging about this topic?'; 'How do students show their understanding of this topic?'" (p. 3). The authors also provide specific suggestions of accessible tools including

- reading systems that read text aloud to students (e.g., VitalSource);
- 3D printing tactile representations of figures and graphs;
- using virtual laboratory simulations (e.g., PhET) to provide extended temporal access to experiments;
- talking calculators;
- following best practices for the physical layout of classrooms;
- training students in how to effectively use the assistive tools.

The literature corpus includes examples of general practices for instructors to engage to support a variety of students. For example, Bustamante *et al.* (2021) provide suggestions (from personal experience and disability literature) about how to support students with attention deficit-hyperactivity disorder (ADHD) in the introductory physics classrooms. The authors suggest that instructors should:

- initiate an open dialogue about students' needs, abilities, and interests;
- scaffold the course content to help students stay on track;
- provide course resources in multiple formats to allow for options in how and when students learn content;
- demonstrate understanding that accommodations promote equity in the class.

While the authors focused their suggestions on how to support students with ADHD, the suggestions apply to all disabled students.

When trying to make a course accessible for disabled students, there are two main non-mutually exclusive paths that can be taken: individual accommodation or inclusive teaching practices. In the accommodation process, disabled students typically request accommodations for their access need(s) via the university's disability services office. This office then writes a letter to instructors describing the approved accommodations and the instructors implement the accommodation(s) for each student. Moon *et al.* (2012) in their report *Accommodating Students with Disabilities in Science, Technology, Engineering, and Mathematics* describe a myriad of accommodations that can be made to support different groups of disabled students. The other main path to include disabled students involves revising or redesigning courses to no longer center able-bodied and able-mindedness and to instead consider learner variation. To do so, instructors and course designers implement inclusive teaching practices that are designed to support a wide range of learners' needs, abilities, and interests. An important difference between this and accommodation is that inclusive teaching practices are implemented for the entire course, while accommodation is done for individuals.

Izzo and Bauer (2015), Lannan *et al.* (2021), Curry *et al.* (2006), and Duerstock and Shingledecker (2014) all suggest the use of universal design for learning (UDL) as a framework for guiding the instantiation of inclusive teaching practices.

1.3.2 Findings for researchers

1.3.2.1 Research on the understanding of disabled students in a particular content area

Two identified articles fit the PER paradigm of researching students' ideas in a particular content area. de Camargo *et al.* (2013) analyzed the interaction of a group of four pre-service physics teachers in Brazil with one blind high school student during lessons on electromagnetism. Their analysis focused on supportive and challenging communication styles, based on "empirical structure," or how "information is materialized, stored, transmitted and perceived" (i.e., visual, tactile, audio-visual, tactile-auditory), and "sensory-semantic structure," or "associative references between meaning and sensory perception" (i.e., inseparable, association, unrelated, and secondary related) (p. 416). The researchers identified the linguistic profiles of 92 communication challenges as independent auditory and visual (i.e., the same information is shown visually and spoken), interdependent audio-visual (i.e., a learner needs to use both sight and hearing to access the information), and interdependent tactile-auditory (i.e., a learner needs to use both sight and touch to access the information). In the electromagnetism context, information that was presented visually included (1) figures demonstrating the processes of charging, electric and magnetic field lines, and circuit and charge configurations; (2) mathematical expressions such as equations, scientific notation, units, and graphs; and (3) values read by measurement instruments. In addition, some information was encoded in inseparable visual representations, such as the characteristic colors of light associated with phenomena. The most frequently identified communication challenge (89/92) was interdependent audio-visual/meaning associated with visual representations, such as "If I have q1 and q2, I have a distance; if I raise it here, it has to decrease there" (p. 417); here, the instructor verbally describes the relationship while visually demonstrating how the parameters are changing. The researchers found that auditory communication and communication styles that combined visual representations with auditory or tactile representations supported communication. The authors argue that tactile-auditory communication supports learning for all students and should be used more frequently in physics instruction.

Bülbül *et al.* (2017) conducted semi-structured interviews with six blind high school students, all girls, about the Force Concept Inventory (FCI). While this article focuses only on high school students, it explores the FCI, which is commonly used in postsecondary STEM and was included in the synthesis. Students' misconceptions about force and motion most frequently fell into four categories: (1) impetus, the belief that an intrinsic force is required to maintain motion; (2) active force, the belief that only active agents, typically living things, exert force; and (3) gravity, the belief that heavier objects fall faster than lighter objects. Thus, the blind students in the study had similar misconceptions about motion as sighted students.

1.3.2.2 Analyzing research-based instructional strategies through a UDL lens

In a recent stream of discipline-based education research (DBER), researchers have analyzed research-based instructional strategies and curricula through the lens of universal design for learning, a framework to support instructions to proactively design instruction that supports the variation in learners' needs, abilities, and interests. Scanlon *et al.* (2018) analyzed popular physics written curricula, including *Tutorials in Introductory Physics, Open Source Tutorials in Physics Sensemaking, Physics by Inquiry, and Next Generation Physics and Everyday Thinking*, to identify examples of UDL-aligned strategies. While these curricula were not intentionally designed to enact UDL-aligned strategies, the researchers found that all four curricula enacted examples of fostering collaboration and community, and supporting planning and strategy development. Multiple curricula also enacted examples of clarifying vocabulary and symbols; and highlighting patterns, critical features, big ideas, and relationships. However, the researchers found few or no examples of practices that supported the spectrum of students' executive function skills (e.g., planning, working memory, time management, and organization), activating or supplying background knowledge, and providing multiple means of engagement. The researchers suggested that curriculum developers consider providing curricular materials in a digital format to allow students to customize the display of information and to access language translation resources; explicitly discuss the use of assistive technologies; explore and incorporate varied means of representation; vary methods of response and navigation; optimize individual choice and autonomy; optimize relevance, value, and authenticity; heighten the salience of goals and objectives; and increase mastery-oriented feedback.

In an extension of this work, Schreffler *et al.* (2017) used an observation protocol based on the UDL guidelines to record the enactment of UDL-aligned strategies in two studio-mode introductory physics courses and two inquiry-based general chemistry laboratories. Observations were conducted before the instructors participated in a year-long faculty learning community about UDL. The observation protocol grouped practices into four categories based on when the practice would likely occur during class: introducing and framing new material, content representation and delivery, expression of understanding, and activity and student engagement. Observations indicated that introducing and framing new material was the area of greatest strength, in terms of implementing UDL-aligned strategies, for the instructors overall; however, there were few examples of instructors assessing background knowledge prior to introducing new material or highlighting what was important for students to learn. Additionally, instructors provided opportunities for collaboration and used "clicker questions" to formatively assess students' understanding, practices aligned with activity, and student engagement. However, within this category, the researchers found few examples of opportunities for students to self-reflect and self-assess. Researchers found few examples of practices aligned with content representation and delivery, such as providing alternatives for students with visual or hearing impairments, and expression of understanding, such as allowing multiple options for how students expressed their understanding.

Google *et al.* (2020) analyzed a case study of "clickers" (personal response devices) to describe how UDL principles can support active learning strategies, such as Peer Instruction. Google *et al.* (2020) state that: (a) Peer Instruction enacts strategies aligned with UDL guideline 7 (provide options for recruiting interest) as students are given the opportunity for individual choice and autonomy by selecting their own response and defending it to peers; (b) questions can optimize relevance, value, and authenticity by using real-world examples; and (c) instructors can minimize threats and distractions by allowing quiet time for students to think before answering and presenting results anonymously. Google *et al.* also argue that Peer Instruction allows students to vary the methods of response, since they answer the question independently, can compare their response to the whole class via the anonymous response distribution and discuss their response with peers. Next, the researchers used a survey to explore instructors' perceptions of whether clickers would provide opportunities to support the UDL principles; their sample consisted of 39 STEM faculty at a university in the southeastern United States. Responses indicated that faculty believed clickers: (1) allow students to monitor their progress and vary methods of response; (2) “support productive feedback, individualized choices, and student autonomy and promote expectations and motivation;” and (3) highlight critical ideas and relationships. On the other hand, faculty were not “aware of how clickers could be used to promote other means of communication… can be used to support access to tools and assistive technology, as well as how clickers allow for multiple media for communication” (p. 961).

1.3.2.3 Research on disabled students’ experiences in postsecondary STEM

Researchers in Project ACCESSS used interviews and interpretative phenomenological analysis to describe the experiences of students who identified with executive function disorders in introductory physics and chemistry courses at a university in the southeastern United States. In an interview study with four participants who identified with ADHD, James *et al.* (2018) found that the lengthy lectures typical of introductory STEM courses did not support students’ learning; students’ learning was better supported when instructors provided breaks during lectures to engage students in clicker questions or student-centered problem-solving. The participants reported that since they were not often actively engaged in class, and they did most of their learning out of class, they described the importance of instructors sharing key dates (e.g., deadlines and exams) and training students to engage with the course materials at the start of the semester. Students also expressed the importance of testing accommodations while at the same time expressing guilt about using those accommodations. In a second study with students who identified with ADHD enrolled in introductory physics courses, researchers found additional support for these challenges (James *et al.*, 2020). Specifically, students reported that the instructors’ time management and organization (e.g., a structured course schedule) could negatively impact their ability to use personal practices essential for course success (e.g., a personal planner). Additionally, while insufficient time on tests created barriers, the extra test time accommodation was sometimes seen as an “unfair” advantage by the students and/or their peers. While one participant with ADHD explained that SCALE-UP courses supported their learning because students have greater autonomy, reducing the impact of becoming distracted, another participant with ADHD[4] found that the physical layout of the SCALE-UP classroom increased distractions because there were other students in all directions.

In a later analysis with nine students who identified with a range of disabilities (five with cognitive impairments, four with emotional/mental health impairments, and one with a visual impairment), researchers identified students’ challenges as related to engaging with the course content and course-related anxiety (James *et al.*, 2019). For example, participants with ADHD reported needing more time than their peers to complete assignments, which could be compounded by STEM content (compared with non-STEM course content) and “flipped class” instructional practices. Additionally, study participants described challenges with misalignment between lectures and labs and content

---

[4] There is not currently commonly accepted impairment-first language for ADHD.

not seeming relevant to their personal interests. The authors explain that "Though these are likely barriers for many students, participants experienced severe consequences, such as being unprepared for assessment, withdrawing from the course, and having anxiety triggered" (p. 260). Seven participants discussed anxiety related to STEM courses, with four participants reporting an increase in frequency and intensity of anxious episodes while taking STEM courses and three participants experienced anxiety related to difficulties preparing for assessments and related to time constraints during assessments. As in the prior study (James *et al.*, 2020) where students with ADHD explained the importance of testing accommodations, students reported that "testing accommodations were critical to the reduction of anxiety" (p. 261). The authors suggest instructional strategies to reduce these barriers based on the UDL framework, including supporting students' studying by highlighting critical features and big ideas with graphic organizers, outline, and weekly quizzes; and supporting students coping with anxiety by promoting external supports (e.g., campus counseling services) and a growth mindset.

Whitney *et al.* (2012) conducted a mixed-methods analysis of students' perspectives on a credit-bearing Learning Community seminar for disabled STEM majors at the University of Southern Maine. Using the social capital theoretical framework (i.e., resources accumulated through relationships that facilitate collective action), the researchers examined what students felt they gained from participation in the seminar. The researchers analyzed responses to pre- and post-seminar surveys of 43 participants, including 11 women and 32 men who predominantly (95%) identified as white; the participants identified with a range of disabilities, including (using the authors' categorization) ADHD or learning disabilities (35%), health-related disabilities (12%), psychiatric/emotional disabilities (11%), autism (7%), orthopedic disabilities (4%), hearing impairment (2%), and traumatic brain injury (2%). Survey data were complimented by a one-hour focus group interview and multiple online discussion forum posts. The researchers found that students gained multiple facets of social capital, including "knowledge, skills, access to resources, and social support" (p. 134). Based on students' responses about their expectations for the seminar, researchers identified high priorities as improved course outcomes, study habits, time management, and career exploration. Moderate priorities included increasing academic support (i.e., academic/non-academic balance, assistive technologies, and STEM career exploration) and social support (i.e., connecting with other students and faculty). Low priorities included improving self-advocacy skills, graduate school exploration/transition, and learning about outside resources and services. Post-survey responses indicated that the seminar did increase social interactions with the program staff, and to a lesser extent with faculty and peers. Additionally, the seminar allowed students to learn about assistive technologies, STEM fields, and local programs and services. However, students still felt only moderately prepared for STEM courses following the seminar.

Jeannis *et al.* (2019) conducted a national survey of the learning barriers and facilitators experienced by students with physical disabilities in instructional science and engineering laboratory settings. The researchers analyzed responses from 107 participants enrolled at 67 unique institutions, who ranged in age from 18 to 68 (57% between 18 and 27), were majority women (65%) and Caucasian (69%), and were in school at the time of the survey (72%). More than 50% of participants reported disabilities related to sitting, kneeling, squatting, or bending, climbing stairs, and/or lifting or carrying objects in their hands or arms. More than 40% of participants reported disabilities related to standing and/or walking. Less frequently reported impacted activities involved fine motor skills and crawling. Around 25% of participants reported barriers in the built environment, such as insufficient signage about accessible entrances. Ramps, elevators, and curb cuts were identified as facilitators in the built environment. Regarding the task execution in the lab, 50% of participants reported that their participation was limited to passive roles, such as notetaking, writing papers, or writing software. More active roles, such as setting up laboratory equipment, were limited due to physical barriers (66%) and time constraints (35%). Course material was the only facilitator commonly discussed for task execution in the lab for this population. Additionally, only 35.5% of participants selected "agree" when asked if "practices were in place to accommodate students with disabilities" (p. 229). However, at least two-thirds of the participants reported positive experiences with instructors (e.g., respectful and inclusive language) and peers (e.g., assistance from peers in completing activities).

Recognizing that much of the research on disabled students' experiences has been conducted in the Global North, Palan (2020) explored the experiences of postsecondary students with visual impairments in India. Through interviews with 29 students, Palan identified four main factors that excluded students from higher education courses in math and science, including exclusion from such courses in earlier education, inadequate support systems, inaccessible teaching practices, and limited job opportunities after graduation.

1.3.2.4 Research on instructors' experiences teaching disabled students

Recognizing that most investigations of UDL-aligned instructional practices had been conducted at four-year institutions, Moriarty (2007) situated her study at three community colleges in Western Massachusetts, as community colleges "enroll the greatest diversity and numbers of students with disabilities" (p. 253). Moriarty collected survey responses from 152 STEM instructors; participants largely identified as white (91%) and were split equally as identifying as men (49%) and women (51%); 36 to 51 years old (40%) and 51 to 65 years old (48%); and teaching full time (57%) and part-time (42%). Moriarty also interviewed 11 of the participants and observed 9 of 11 in the classroom. Many respondents (42%) indicated that the majority of each class period is spent in the traditional lecture format, while varied presentation strategies showed the highest reported use (3.93/5 adjusted mean). Instructors predominantly reported using traditional assessments, including exams (89%) and projects (56%), and less frequently reported using papers (37%) or portfolios (19%). On the inclusive mindset scale, 78% of respondents indicated "they agree or strongly agree that they are receptive to making changes to accommodate students with disabilities, and 75% agree that students with disabilities are capable of learning the material in their class. Respondents also agree that they try to match their teaching styles to accommodate students" learning needs (74%), and they agree that they continually look for better ways to teach and are open to new forms of instruction (88%)" (p. 257). Additionally, there was a slight trend toward instructors indicating that they did have the time and resources to develop new teaching approaches. Inclusive mindset and technology comfort level were significantly correlated with many inclusive instructional practices; these findings were supported by regression analysis, which also identified time for instructional development to be predictive of varied presentation strategies, interactive learning, student engagement, and pedagogical variety. Based on interviews and observations, Moriarty stated "It appears that for the most part, instructors are aware of diversity and the need for inclusion and attempt to teach in ways that reach a diverse population of students. Nevertheless, findings related to the use of materials and technologies in the classroom suggest that improvement is needed in the area of accessibility" (p. 260). Barriers to using more inclusive teaching methods were dominated by financial and institutional demand, such as high teaching load and lack of time to develop new methods. Overall, Moriarty concluded that the findings indicated that community college faculty "appear more knowledgeable about pedagogical practices than what has been reported in previous literature about four-year faculty" (p. 264).

Shmulsky *et al.* (2018) analyzed interviews with 12 STEM instructors at "a liberal arts college that serves students who learn differently" (about one-third of students had a documented autism spectrum disorder diagnosis) to identify instructors' perceptions of strengths and challenges for autistic students and general personal traits needed for success in STEM fields. Participants taught courses in biology, chemistry, computer science, physics, and mathematics. The interviewed instructors emphasized the variability in both profile and effective teaching strategies for autistic students; for example, some autistic students were viewed as concrete thinkers, while others had a strong ability to think abstractly. Thus, Shmulsky, Gobbo, and Bower qualify their findings with the essential recognition that all students are unique. Participants reported that autistic individuals tend to have STEM-relevant strengths related to attending to detail, following complex directions, and recognizing and using patterns. On the other hand, participants reported common STEM-relevant challenges related to expressing frustration, social interaction (e.g., over- or under-participating in a group discussion), and rigidity/inflexibility. Participants described how rigidity and inflexibility could be assets in STEM fields, such as supporting precision necessary for measuring chemicals, solving lengthy math problems, and debugging computer code, as well as persistence in the STEM major. The researchers conclude "Teaching implications of this research include the importance of developing and using strategies to support social interaction and critical thinking…[and] finding practical ways to engage students' strengths" (p. 53).

Gokool-Baurhoo and Asghar (2019) interviewed 18 instructors, including five physics instructors, who had experience teaching students with a learning disability (LD) at an English-language CEGEP (publicly funded college) in Quebec, Canada. Researchers found that half of the instructors reported a lack of knowledge and skills to teach science to students with learning disability due to difficulty identifying relevant challenges and creating accessible science instruction. This was identified as a "second-order barrier," indicating this barrier is internal to the teacher. Instructors also reported insufficient support in working with disabled students, including not knowing each student's specific disabiliti(es) and lack of training and professional development. 75% of instructors discussed "certain negative attitudes and difficult behaviours including: reluctance to share information about their LD and seek academic support from their teachers; a persistent lack of engagement with science; and difficult and anxiety-ridden behaviours" commonly exhibited by students with learning disabilities that could lead instructors to have difficulty with establishing relationships with students with learning disabilities (p. 23). Gokool-Baurhoo and Asghar identified these challenges as "first-order barriers," meaning they are "external to the teacher and stem mostly from the environment" (p. 19). Gokool-Baurhoo and Asghar suggest that disability service office personnel "emphasize to these students that their teachers might be willing to further accommodate their academic needs, should they choose to disclose their disabilities" (p. 25). Additionally, the researchers call for hands-on, authentic professional development in supporting students with learning disabilities in science courses.

Based on views like those expressed in Gokoll-Baurhoo and Ashjar's study, Scanlon and Chini (2018) designed a framework for proactively considering how specific learning experiences may privilege and simultaneously tax particular "dimensions of ability." Using literature from disability studies, education, medicine, social science, psychology, technology, and governmental organizations, Scanlon and Chini identified six dimensions of ability along which individuals vary (updated from original paper): physical/mobility, health, cognitive, visual, hearing, and emotional/mental health. Rather than categorizing individual students, Scanlon and Chini invite "instructors, curriculum developers, and researchers to apply the framework to the curricular materials and learning environment and ask questions such as 'What load does this activity put on each dimension?' and 'Overall, does my course frequently place a high load on certain dimensions in a way that privileges certain abilities?'" (p. 2). Scanlon and Chini provide examples of using the framework to consider the expected load on each dimension for popular instructional activities. For example, they state that traditional lecture and small group problem-solving would both load high on the hearing dimension to engage in verbal communication (e.g., listen to the instructor or peers). Individual clicker questions and hands-on activities do not necessarily require students to listen to someone else, so these activities load lower on the hearing dimension. This tool is intended to allow instructors to identify and plan options for potential challenges without knowing individual students' diagnoses.

Scanlon and Chini (2020) also modified and piloted a survey of physics instructors' views about supporting learner variation. Starting with the cross-disciplinary Inclusive Teaching Strategies Inventory (ITSI; and Lombardi *et al.*, 2011), which assesses instructors' beliefs and actions related to disability and supporting disabled students, Scanlon and Chini made modifications based on a prior pilot administration of the ITSI. In the first pilot administration, physics graduate students and physics and chemistry faculty took the ITSI and shared their thoughts; they expressed challenges, such as wanting to indicate the population of students they had in mind for each prompt and not viewing some of the prompted instructional practices (e.g., discussion board prompts) as relevant to postsecondary physics instruction. Scanlon and Chini also conducted interviews with physics instructors and discussed changes with the ITSI developer; modifications were made to allow respondents to mark the population (i.e., no students, only students with disabilities, students who need it, or all students) and to clarify instructional practices and/or make them more relevant to typical physics instruction.

They then piloted the modified survey at two in-person professional conferences and collected 13 validated responses, including eight men, four women, and one non-binary person; four students, eight post-secondary faculty, and one industry member; ten participants who stated they had worked with or taught disabled students, either with personal contact with a disabled person, one who personally identified with disability, or two who stated they had no personal experience with disabled people. The pilot probed the population of participants considered within the two student categories: "only students with disabilities" and "students who need it." Participants were given four options: "I. Students registered with the disability service office on campus, II. Students not registered with disability services office but who have diagnosed disability, III. Students who identify with disability (i.e., undiagnosed), and IV. Other, please specify. Five participants selected the only option I, two selected only option III, four selected I and II, and two selected I, II, and III." (p. 5). Thus, most instructors selected students registered with the disability service office for the students with disabilities category and the researchers suggest that the two who did not may have been confused by the wording of option III. However, the researchers also point out meaningful variation in who respondents were included in "only students with disabilities." Scanlon and Chini asked respondents who they included in "students who need it" as an open-ended prompt. Responses included students who express a need to the professor or self-identify as needing accommodations (6/13 participants); students who have extenuating circumstances outside the classroom, sometimes with an emphasis on a "valid" excuse (3/13); and students whose learning would be significantly impacted by accommodation. The researchers caution that "If an instructor does not have the same types of life experiences (such as disability, family, or financial) as the student requesting accommodation (which is unlikely), then the instructor may find it difficult to determine what is and is not a 'valid' excuse" (p. 5).

Research in physics education has just begun to explore the experiences of disabled learners. It is essential that research centers on the knowledge, skills, and experiences of these disabled learners, in contrast to prior work that has centered instructors' labor in teaching disabled students. Researchers should also explore how students and instructors' multiple identities shape the experiences of disabled physics learners.

1.4 DISCUSSION

1.4.1 Critiques of the literature

Many of the sources in the literature corpus contain an introductory framing that is problematic; specifically, the articles cite the increase in representation of disabled students in higher education and frame this as a burden on instructors. For example, Gough (1978) states

> "All too often the first inkling a science teacher has that a visually impaired student is a member of his or her class comes when they meet face to face. If you are that teacher, you may experience

> shock, frustration, anger, or any number of uncomfortable emotions. How can you cope with yet another "problem," you may wonder? How can you provide for that student's safety? How much extra time will you have to spend? What should the student be allowed to do, and what not" (p. 34).

More recently published articles include similar framings. An alternate and more positive framing is that all students, regardless of their disability status, have a variety of needs, abilities, and interests (Scanlon and Chini, 2018). Therefore, instructors should plan for learner variation in their courses from the start, and the presence of disabled learners in physics courses is not a surprise or aberration but instead an expected student variation.

Additionally, many of the sources in the literature corpus focus on visually impaired students. Specifically, of the 66 sources that focus on a particular type of impairment, 62% (33 sources) focus on supporting visually impaired students, while 9.4% focused on cognitively impaired students, 9.4% on physically/mobility impaired students, and 7.5% on hearing impaired students. In the literature corpus, there were no articles that focused on health impaired nor emotional/mental health-impaired students. This is concerning in the context of a recent study from Fall 2020, which shows the representations of disabled students, many of whom were enrolled in emergency-remote courses, in U.S. higher education (Scanlon *et al*., 2021). In this study, 61.5% of all disabled students had cognitive impairments, 41.2% emotional/mental health impairments, 17.6% health impairments, 2.0% hearing impairments, 1.4% physical/mobility impairments, and 1.4% visual impairments. Numerous participants were identified with multiple impairments. This shows that the representation of literature focused on supporting visually impaired students is disproportioned compared with the representation of visually impaired students in physics courses. Relatedly, none of the sources highlighted intersectional identities or discussed how people's intersectional identities may affect their experiences in physics learning environments.

Another trend in the literature corpus is that many recent articles include recommendations that were published decades ago. For example, Henderson (1965) and Holt *et al.* (2019) suggest the use of raised lines as an alternative representation of diagrams and graphs. Multiple sources suggest the use of Braille labels (even though recent trends show decreased Braille-literacy amongst blind and low-vision people; Kleege, 2006) and tactile metersticks and other measurement devices. This repetition could indicate that there has not been much uptake of these suggestions by practitioners and/or that recent authors are unaware of previous articles with similar suggestions. Relatedly, many articles do not catalog or use existing literature, campus, or community resources for developing their own. For example, many sources do not mention their local office of disability services, which are common in U.S. institutions and are the main mechanism by which disabled students receive accommodations to meet their access and inclusion needs.

Many of the practitioner-focused sources center around addressing concrete access issues, while little attention has been paid to changing instructors' beliefs and/or mindsets. For example, 24 sources focused on support access and inclusion in the laboratory setting and most sources focused solely on providing access to laboratory equipment and experiments rather than expanding to discuss broader culture and climate topics. Additionally, few articles have discussed issues related to instructor beliefs, while prior research shows that STEM faculty hold more negative beliefs about disability than their counterparts in other disciplines (Rao, 2004). Finally, the sources identified do not form a well-connected network because they infrequently cite each other.

1.4.2 Implications for practitioners

In addition to the concrete suggestions included in the findings section, instructors should focus their efforts toward the following: (a) identifying barriers in their courses; (b) identifying inclusive instructional strategies that can lower and/or eradicate the barriers; (c) implementing the identified strategies; and (d) assessing the impact of the new strategies. As a first step, instructors should critically examine their courses along accessibility and inclusivity lines. As people all have different needs, abilities, interests, and lived experiences, instructors should include a wide range of stakeholders, including disabled and non-disabled students, in their course examinations. To identify barriers to access and inclusion, instructors should consider the different types of abilities, as described in Scanlon and Chini (2018). When identifying barriers to access and inclusion, instructors should consider a variety of students who may have strengths and limitations along each dimension of ability. If a course is composed of instructional strategies that all load high on a dimension of ability, then the course continually privileges students with strengths along that dimension while simultaneously taxing students with limitations along that dimension. This process can be used to identify barriers. Chini and Scanlon (2021) provide examples of this process in their AAAS blogpost.

Once barriers to access and participation have been identified, instructors should identify inclusive instructional strategies that can be used in their courses to reduce or eliminate these barriers. Universal design for learning (UDL; and CAST, 2018) is a design framework that supports instructors in proactively creating learning environments that support the broadest range of students without the need for specialized modifications. Recent research also includes suggestions of inclusive teaching strategies specific to physics (Scanlon *et al.*, 2018; Bustamante *et al.*, 2021; Chini *et al.*, 2021; and Lannan *et al.*, 2021). Using these tools, instructors should plan for a wide range of students' needs, abilities, and interests through iterative improvement. Next, instructors should implement inclusive teaching strategies in their courses.

Previous research in chemistry education shows that when sighted chemistry instructors worked to create alternative representations of gas law topics, the visually impaired students who used the developed tools were overwhelmed and found the tools difficult to use with other assistive technologies (Harshman *et al.*, 2013). In order to value the lived experience and knowledge of disabled students about their own body and needs, instructors should partner with disabled students in the development of access and inclusion strategies. (Note: This does not mean that instructors should expect disabled students to explain their lived experiences, inform instructors about disability background knowledge, and be the sole advocate for themselves and their needs. Instructors should carefully consider how to ethically involve students who have likely experienced ableism and disablism in their lives and are traditionally marginalized.) Similarly, Seymour and Hunter (1997) state: “It is important for those seeking to improve the higher education chances of this group of students [disabled students] and wondering where to place their emphasis, to have clear directions from the students themselves about what they need the most” (p. 185).

Prior research shows that STEM faculty hold more negative attitudes and beliefs about disability than do their colleagues in other disciplines (Rao, 2004), and that physics faculty lack knowledge about disability diagnoses and hold beliefs about the viability of physics careers that gatekeep who are supported to join the physics professional communities (Scanlon *et al.*, 2020; and Oleynik *et al.*, 2021). Therefore, instructors should engage in professional development to gain knowledge and fluency with disability topics and should focus on sensemaking about disability to shift their mindsets to be more positive.

1.4.3 Implications for researchers

While there are numerous practitioner-focused articles on disability, there is a dearth of education studies on best practices for supporting disabled students in physics courses. Researchers should engage in additional research to: identify inclusive teaching strategies to support disabled students, investigate the unique experiences of disabled students in physics programs, and investigate the intersectional experiences of students (e.g., experiences of disabled students of color, disabled women, disabled LGBTQ + people). Additionally, much of the literature reviewed in this chapter centered on visual, hearing, and/or physical/mobility impairments. Researchers should conduct research with and for emotional/mental health, cognitive, and health-impaired students, especially since such impairments have a high prevalence in the physics community.

While identifying solutions to concrete access needs is important, many of the practitioner-focused sources did not include efficacy studies of the impacts of the solutions. Researchers should take up suggestions of solutions to access needs and investigate their impact on diverse populations of physics learners. Researchers should also move past solely focusing on access needs and investigate the broader ecosystem of higher education and systems of ableism and disablism in physics communities. Seymour and Hunter (1997) state: “in order for the potential of students with disabilities to be fully realized, and the risk of losing good students minimized, priority should now be given to changing [STEM] faculty attitudes” (p. 185). In recognition that ableism is systemic and not merely personal, researchers should investigate the systemic changes necessary to make available time and resources for instructional development. A common phrase in the disability justice realm is “nothing about us without us.” Researchers should ethically work with (not just for) disabled students to provide avenues for disabled students to share their experience and expertise about their own access needs. This also pushes back against paternalism endemic in disability advocacy.

In working directly with disabled populations, researchers should remain attentive to community identities, which may contrast with academic literature in certain disciplines. For example, person-first language has been centered in research on K-12 education, while many disabled individuals have shifted toward impairment-first language. Researchers should intentionally choose the language they use and explain their choices.

Education researchers who do not focus their work on disability also play a role. Most physics education research papers that include students do not include disability as a demographic variable and/or category. Researchers should judiciously collect disability status information and report these findings in their studies. Disabled students are present in physics courses and not including their identities in studies is erasing their existence from physics, thereby perpetuating the notion that disabled students are aberrations in physics courses. Additionally, when studies identify an interesting trend related to disability, the authors should report this trend (as an example of this practice see Gandhi *et al.*, 2016). Everyone has a role to play in dismantling ableism in physics.

1.5 FUTURE DIRECTIONS

Looking toward the future, there are several essential directions for researchers in physics education to explore. First, researchers should attend to the intersectionality of individual identities and the varying impact of disability across other dimensions of identity, such as gender, race/ethnicity, national origin, sexual identity, and combinations of disabilities. Hawley *et al.* (2013) reviewed the literature on underrepresented minority disabled students and found “At each transitional phase (elementary school to middle school to high school to post-secondary school), large numbers of URM/SWD individuals are ‘redirected’ from STEM long-term goals as well as the educational, social, and psychological experiences necessary to achieve them” due to “several systemic and serious impediments of an educational, psychological, economic, and attitudinal nature that in the aggregate serves to severely limit the numbers of STEM candidates in higher education” (p. 94). For example, African American students are overrepresented in U.S. special education via educational diagnoses of intellectual disability and/or emotional disturbance; Hispanic students are overrepresented via education diagnoses of hearing impairments and learning disabilities. Additionally, gender, family income, and home language also increase a student’s chances of being placed in special education, with disproportionate representation of boys, families living in poverty, and families who speak a language other than English at home. Students in special education are less likely to take high school course work that would prepare them for postsecondary STEM majors. Da Silva Cardoso *et al.* (2013) used hierarchical regression analysis to identify significant predictors of STEM goal persistence for 115 URM disabled students and found that gender, advanced placement (AP) classes, father’s educational level, academic milestone self-efficacy, and STEM interest accounted for 57% of the variance in STEM persistence. Additional analysis demonstrated that the Social-Cognitive Career Theory “provides useful guidance for designing postsecondary education interventions for minority disabled students in STEM education to help crystalize their career interest and increase goal persistence" (Dutta *et al.*, 2015, p. 159). Coleman (2017) conducted a case study analysis of four women with sensory and mobility impairments in STEM careers and found that participating women "felt more gender-based barriers during STEM education and in their career than barriers related to their disability" (p. 150). Researchers should continue to explore how individuals' multiple identities intersect with disability in the physics community.

Researchers should also zoom out from medical-model-aligned individual classroom accommodations and consider the broader ecosystem of STEM education and careers. Earlier work often called for training disabled students to self-advocate. Pfeifer *et al.* (2020) conducted interviews with 25 STEM majors who received accommodations for ADHD and specific learning disabilities and revised a generic conceptual model of self-advocacy for disabled individuals to focus on ADHD and specific learning disabilities in undergraduate STEM courses. Test's original conceptual framework for self-advocacy includes four components: knowledge of self, including strengths and weaknesses as a student and as a disabled person; knowledge of rights, including laws and policies relevant to the accommodation process in college; communication, including acceptable communication behaviors; and, optionally, leadership, including awareness of individuals responsibilities to advocating on behalf of others (Test *et al.*, 2005). Emergent components of self-advocacy based on the experiences of students with ADHD and/or SLD in STEM courses included knowledge of accommodations and the process of attaining them; knowledge of the influence of STEM learning contexts on accommodation needs; "filling gaps," or "participant actions taken to overcome limitations in formal accommodations or instructional supports" such as creating a collaborative Google Docs for notetaking in response to poor quality provided notes or finding tutors when course instructors are not approachable. Additionally, the researchers identified emergent beliefs that impacted participants' self-advocacy, such as agency and view of disability, with students who identified positive aspects of their disability more likely to access accommodations. The researchers highlight the need for more research to support or refute these proposed components of self-advocacy for undergraduate STEM. Researchers should also consider how the postsecondary STEM education system can change to lower the barrier to and/or reduce the need for self-advocacy, such as through broad implementation of inclusive teaching practices.

It is also important to zoom out from classroom accommodations to consider how disabled individuals are able to participate in the broader physics community, such as physics research. Several studies describe the integration of Deaf and hard-of-hearing (HH) students in STEM research. Pagano *et al.* (2015) describe a research experience for undergraduates (REU) program at an institution with an integrated school for Deaf students. They repeat that concerns about involving Deaf/HH students in research labs, such as safety, are typically due to faculty lack of knowledge rather than true safety issues. Smith *et al.* (2016c) summarize useful safety strategies, such as instituting an emergency notification system. The researchers describe strategies for successful undergraduate research experiences, including faculty mentors engaging in specialized American Sign Language (ASL) classes focused on "core scientific terminology and laboratory and field safety" as "even a small base knowledge of key signs can increase communication and be very effective in strengthening relationships between students, peers, and mentors" (p. 153). At the same time, Pagano *et al.* (2015) describe the importance of enabling Deaf/HH students to take leadership in communication and to teach the research group essential signs for the research environment. Communication facilitation, via ASL interpreters, CART specialists, texting features on cell phones, or video remote interpreting, are frequently needed in research labs and conferences. Gehret *et al.* (2017) surveyed Deaf/HH students and their research mentors and identified challenges such as students feeling socially isolated and missing out on "ambient knowledge" when communication facilitation was not available in the lab. Ott *et al.* (2020) interviewed ASL interpreters who had worked with teams of one Deaf and two hearing students during six-week internships and identified unique challenges for the interpreters in the research environment, such as deciding when and how to interpret. Researchers should continue this line of work and expand it to other impairments and STEM disciplines.

Lillywhite and Wolbring (2019) reviewed the literature for examples of research on undergraduate disabled students as "knowledge producers, including as researchers" (p. 1). They identified only 15 relevant articles and did not find any studies that investigated how undergraduate students chose a research topic or were recruited to join research projects. Lillywhite and Wolbring call for disability studies and STEM education researchers to explore the role of undergraduate disabled students as knowledge producers and researchers.

Continuing to zoom out from academic learning, Pacheco (2014) investigated career choice and participation by STEM professionals and graduate students with sensory and orthopedic disabilities through interviews with 18 participants. Findings suggested that social persuasion played an important role in self-efficacy for the participants. Additionally, assistive technology was critical for participation in STEM, and barriers to participation included gatekeepers' limiting perceptions and lack of knowledge about relevant assistive technologies.

After a long history of positioning disabled individuals as a burden in the physics community, it is time for physics education research to center the knowledge, skills, and experiences of disabled individuals and to identify systemic change that will support full participation of disabled individuals in physics classrooms, laboratories, research experiences, and the broader community.

REFERENCES

Annamma, S. A. *et al.*, Race Ethn. Educ. **16**(1), 1–31 (2013).
Azevedo, A. D. and Santos, A. C. F., Phys. Educ. **49**(4), 383–386 (2014a).
Azevedo, A. C. and Santos, A. C. F., Rev. Bras. Ensino Fís. **36**(4), 01– (2014b).
Baker, K. Q. *et al.*, J. Postsecond. Educ. Disabil. **25**(4), 309–329 (2012).
Baughman, Jr., J. and Zollman, D., Phys. Teach. 15(6), 339–342 (1977).
Bernhard, K. and Bernhard, J., presented at the International Conference Practical Work in Science Education, Copenhagen, Denmark, 20–23 May 1998.
Bernhard, K. and Bernhard, J., Phys. Teach. **37**(9), 555–556 (1999).
Brazier, M. *et al.*, J. Coll. Sci. Teach. **30**(2), 114 (2000).

Bülbül, M. S., ERIC No. ED516639 (2009).

Bülbül, M. Ş. *et al.*, Phys. Educ. **48**(3), 275–276 (2013).

Bülbül, M. Ş. *et al.*, Eur. J. Phys. Educ. **6**(3), 20–31 (2017).

Burgstahler, S., Inform. Technol. Disabil. **1**(4), 4 (1994).

Bustamante, C. *et al.*, Phys. Teach. **59**(573), 573–576 (2021).

Carver, T. R., Phys. Teach. **5**(2), 61–65 (1967).

CAST, Universal Design for Learning Guidelines version 2.2 (2018), retrieved from http://udlguidelines.cast.org.

Cetera, M. M., J. Coll. Sci. Teach. **12**(6), 384–393 (1983).

Chini, J. K. and Scanlon, E., AAAS IUSE Blog (2021).

Chini, J. J. *et al.*, *PERC Proceedings*, edited by M. B. Bennett *et al.* (Virtual Conference, August 4–5, 2021), pp 99–104.

Cole, R. A. and Slavin, A. J., J. Vis. Impair. Blind. **107**(4), 311–315 (2013).

Coleman, S. B., *Doctoral dissertation* (Drake University, 2017).

Curry, C. *et al.*, Sci. Teach. **73**(3), 32 (2006).

da Silva Cardoso, E. *et al.*, Rehabil. Res. Policy Educ. **27**(4), 271 (2013).

de Azevedo, A. C. *et al.*, Phys. Educ. **50**(1), 15–18 (2015).

De Camargo, E. P. *et al.*, J. Emerg. Trends Educ. Res. Policy Stud. **4**(3), 413–423 (2013).

Dickman, A. G. *et al.*, Phys. Educ. **49**(5), 526–531 (2014).

Dounas-Frazer, D. R. *et al.*, arXiv:2202.00816 (2022).

Duerstock, B. S. and Shingledecker, C. A., *The American Association for the Advancement of Science* (AAAS, 2014).

Dutta, A. *et al.*, J. Vocat. Rehabil. **43**(2), 159–167 (2015).

ESCAP, UN, "Disability in Asia and the Pacific," Midpoint Review (2017); see https://repository.unescap.org/handle/20.500.12870/542.

Evans, N. J. *et al.*, *Disability in Higher Education: A Social Justice Approach* (John Wiley & Sons, 2017).

Farrell, B. *et al.*, NSF 2001 Report: Engineering Senior Design Projects to Aid Persons with Disabilities (Wayne State University, 2001), pp. 308–309; see, http://nsf-pad.bme.uconn.edu/2001/Wayne%20State%20University.pdf.

Fell, E. V. *et al.*, *Research Anthology on Physical and Intellectual Disabilities in an Inclusive Society* (IGI Global, 1985).

Flaherty, M. and Roussy, A., Edu. Law J. **24**(1), 1–23 (2014).

Frinks, R. M., Phys. Teach. **21**(8), 536–537 (1983).

Gandhi, P. R. *et al.*, Am. J. Phys. **84**(9), 696–703 (2016).

Gehret, A. U. *et al.*, J. Sci. Educ. Stud. Disabil. **20**(1), 20–35 (2017).

Gokool-Baurhoo, N. and Asghar, A., Teach. Teach. Educ. **79**(12), 17–27 (2019).

Goncalves, A. M. B. *et al.*, Phys. Educ. **52**(5), 053002 (2017).

Google, A. N. *et al.*, *Active Learning in College Science* (Springer, Cham, 2020), pp. 953–964.

Gough, E. R., Sci. Teach. **45**(9), 34–35 (1978).

Harshman, J. *et al.*, J. Chem. Educ. **90**(6), 710–716 (2013).

Hawley, C. E. *et al.*, J. Vocat. Rehabil. **39**(3), 193–204 (2013).

Henderson, D. R., ERIC No. ED011155 (1965).

Holt, M. *et al.*, Phys. Teach. **57**(2), 94–98 (2019).

Institute of Physics, see https://www.iop.org/publications/iop/2008/page_42867.html for (2008).

Izzo, M. V. and Bauer, W. M., Univers. Access. Inf. Soc. **14**(1), 17–27 (2015).

James, W. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020111 (2020).

James, W. *et al.*, *2017 Physics Education Research Conference Proceedings* (AAPT, 2018), pp. 192–195.

James, W. *et al.*, *2019 PERC Proceedings*, 24–25 July 2019, edited by Y. Cao *et al.* (PERC, Provo, UT, 2019).

Jain-Link, P. and Kennedy, J. T., Harv. Bus. Rev. (2019).

Jeannis, H. *et al.*, Disabil. Rehabil.: Assist. Technol. **15**, 225–237 (2019).

Kirkup, G. *et al.*, *Women and Men in Science, Engineering and Technology: The UK Statistics Guide 2010* (UK Resources Centre for Women in Science and Technology, Bradford, 2010).

Kleege, G., J. Vis. Cult. **5**, 209–218 (2006).

Kolne, K. and Lindsay, S., J. Occup. Sci. **27**(4), 525–546 (2020).

Kouroupetroglou, G. and Kacorri, H., AIP Conf. Proc. **1203**(1), 1308–1313 (2010).

Lannan, A. *et al.*, Phys. Teach. **59**(3), 192–195 (2021).

Lang, H. G., Phys. Teach. **11**(9), 527–531 (1973).

Lang, H. G., Phys. Teach. **19**(4), 248–249 (1981).

Leddy, M. H., J. Spec. Educ. Technol. **25**(3), 3–8 (2010).

Lee, A., Career Develop. Exceptional Individuals **34**(2), 72–82 (2011).

Lee, A., Int. J. Disabil. Develop. Educ. **69**, 1295–1312 (2022).

Lewis, M. L., *Doctoral dissertation* (Auburn University, 1998). Dissertation Abstracts International, 59, 08-A (1998).

Leyser, Y. *et al.*, J. Postsecond. Educ. Disabil. **13**(3), 5–19 (1998).

Li, Y. *et al.*, Int. J. STEM Educ. **7**(1), 1 (2020).

Lillywhite, A. and Wolbring, G., Educ. Sci. **9**(4), 259 (2019).

Lombardi, A. R. *et al.*, J. Divers. High. Educ. **4**(4), 250–261 (2011).

Mansoor, A. *et al.*, Disabil. Rehabil.: Assist. Technol. **5**(2), 143–152 (2009).

Moon, N. W. *et al.*, *Accommodating Students with Disabilities in Science, Technology, Engineering, and Mathematics (STEM)* (Center for Assistive Technology and Environmental Access, Georgia Institute of Technology, Atlanta, GA, 2012), pp. 8–21

Moore, E. B. and Perkins, K. K., *Cyber-Physical Laboratories in Engineering and Science Education* (Springer, Cham, 2018), pp. 141–162.

Moore, E. B. *et al.*, *International Conference on Universal Access in Human-Computer Interaction* (Springer, Cham, 2018), Vol. 10907, pp. 385–400.

Moriarty, M. A., Equity. Excell. Educ. **40**(3), 252–265 (2007).

National Center for Science and Engineering Statistics, Women, minorities, and persons with disabilities in science and engineering: 2021. Special Report NSF 21-321 (National Science Foundation, Alexandria, VA, 2021), see https://ncses.nsf.gov/wmpd.

Negrete, O. *et al.*, Eur. J. Phys. **41**(3), 035704 (2020).

Norman, K. *et al.*, Sci. Educ. **82**(2), 127–146 (1998).

Oleynik, D. P. *et al.*, *PERC Proceedings*, edited by M. B. Bennett *et al.* (Virtual Conference, August 4–5, 2021).

Ott, L. E. *et al.*, J. Microbiol. Biol. Educ. **21**(1), 20 (2020).

Pacheco, H. A., *Choice and Participation of Career by STEM Professionals with Sensory and Orthopedic Disabilities and the Roles of Assistive Technologies* (Arizona State University, 2014).

Pagano, T. *et al.*, Educ. Sci. **5**(2), 146–165 (2015).

Palan, R., Disabil. Soc. **36**(2), 202–225 (2020).

Parry, M. *et al.*, Phys. Teach. **35**(8), 470–474 (1997).

Pfeifer, M. A. *et al.*, Int. J. STEM Educ. **7**(1), 1– (2020).

Prema, D. and Dhand, R., Can. J. Disabil. Stud. **8**(3), 121–141 (2019).

Rao, S., Coll. Stud. J. **38**(2), 191–198 (2004).

Reed, M. *et al.*, Can. J. High. Educ. **33**(2), 27–56 (2003).

Scanlon, E. M. and Chini, J. J., *2018 PERC Proceedings*, edited by A. Traxler *et al.* (Washington, DC, August 1–2, 2018).

Scanlon, E. M. and Chini, J. J., *2019 Physics Education Research Conference* (AAPT, 2020).

Scanlon, E. M. *et al.*, *2020 PERC Proceedings*, edited by S. Wolf *et al.* (Virtual Conference, July 22–23, 2020).

Scanlon, E. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(2), 020101 (2018).

Scanlon, E. M. *et al.*, *2021 PERC Proceedings*, edited by M. B. Bennett *et al.* (Virtual Conference, August 4–5, 2021).

Schoen, E. *et al.*, Coll. Stud. J. **21**(2), 190–193 (1986).

Schreffler, J. *et al.*, Int. J. STEM Educ. **6**(1), 1 (2019).

Schreffler, J. *et al.*, *2017 PERC Proceedings*, edited by L. Ding *et al.* (Cincinnati, OH, July 26–27, 2017), pp. 360–363.

Science Association for Persons with Disabilities, *Bibliography of Publications Relating to the Teaching of Science to Students with Disabilities* (ERIC Clearinghouse, Cedar Falls, IA, 1994).

Sevilla, J. *et al.*, Phys. Educ. **26**(4), 227–230 (1991).

Seymour, E. and Hewitt, N. M., *Talking About Leaving: Why Undergraduates Leave the Sciences* (Westview Press, 1997).

Shakespeare, T., Disabil. Stud. Read. **2**, 197–204 (2006).

Sherbin, L. *et al.*, Disabilities and Inclusion (Global and US Findings) (2017).

Shmulsky, S. *et al.*, J. STEM Teach. Educ. 53(2), 4 (2018).

Sica, M. G., ERIC No. ED220280 (1982).

Skinner, M., Int. J. Spec. Educ. **22**(2), 32–45 (2007); available at https://files.eric.ed.gov/fulltext/EJ814486.pdf.

Slavin, A. J., Can. J. Phys. **93**(1), 1–2 (2014).

Smith, T. L. *et al.*, *International Conference on Universal Access in Human-Computer Interaction* (Springer, Cham, 2016a), Vol. 9739, pp. 147–158.

Smith, T. L. *et al.*, *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (ACM, 2016b), pp. 319–320.

Smith, T. L. *et al.*, J. Technol. Persons Disabil. 225–238 (2017).

Smith, S. B. *et al.*, J. Chem. Health Saf. **23**(1), 24–31 (2016c).

Stewart, T., Phys. Teach. **18**(4), 291–293 (1980).

Sukhai, M. A. and Mohler, C. E., *Creating a Culture of Accessibility in the Sciences* (Academic Press, 2016).

Supalo, C. A. *et al.*, J. Sci. Educ. Stud. Disabil. **12**(1), 27–32 (2007).

Test, D. W. *et al.*, Remed. Special Edu. **26**(1), 43–54 (2005).

Thompson, D. M., *International Conference on Technology and Persons with Disabilities (CSUN)* (Springer, 2005).

Thompson, A. R. *et al.*, Rehabil. Couns. Bull. **40**(3), 166–180 (1997).

Thurston, L. P. *et al.*, J. Postsecond. Educ. Disabil. **30**(1), 49–60 (2017).

Tomac, M. *et al.*, Phys. Teach. **54**(5), 285–287 (2016).

Tomlinson, B. J. *et al.*, J. Technol. Persons Disabil. (2019).

Tomlinson, B. J. *et al.*, J. Technol. Persons Disabil. **6**, 202–218 (2018).

Torres, J. P. and Mendes, E. G., Phys. Teach. **55**(7), 398–400 (2017).

Traxler, A. and Blue, J., *Physics Education and Gender* (Springer, Cham, 2020), Vol. 19, pp. 129–152.

Truncale, N. P. and Graham, M. T., Phys. Teach. **52**(2), 76–79 (2014).

United Nations, *Disability Laws and Acts by Country/Area* (Department of Economic and Social Affairs Disability, n.d.); see https://www.un.org/development/desa/disabilities/disability-laws-and-acts-by-country-area.html.

United Nations Convention on the Rights of Persons with Disabilities, December 13, 2006, see https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html.

Van Domelen, D. J., Phys. Teach. **37**(8), 500–501 (1999).
Vongsawad, C. T. *et al.*, Phys. Teach. **54**(6), 369–371 (2016).
Weems, B., Phys. Teach. **15**(6), 333–338 (1977).
Whitney, J. *et al.*, J. Postsecond. Educ. Disabil. **25**(2), 131–144 (2012).
Windelborn, A. F., Phys. Teach. **37**(6), 366–367 (1999).
Winters, R. M. *et al.*, Ergon. Des. **27**(1), 5–10 (2018).
World Health Organization, *International Classification of Functioning, Disability and Health (ICF)* (WHO, Geneva, 2001).
World Health Organization, *World Report on Disability* (WHO, Geneva, 2011).
World Wide Web Consortium, see https://www.w3.org/WAI/policies for (2019).
Zhang, D. *et al.*, Remedial Spec. Educ **31**(4), 276–286 (2010).

CHAPTER 2

FRAMEWORK FOR AND REVIEW OF RESEARCH ON ASSESSING AND IMPROVING EQUITY AND INCLUSION IN UNDERGRADUATE PHYSICS LEARNING ENVIRONMENTS

Sonja Cwik and Chandralekha Singh

Cwik, S. and Singh, C., "Framework for and review of research on assessing and improving equity and inclusion in undergraduate physics learning environments," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 2-1–2-26.

2.1 INTRODUCTION AND FRAMEWORK

Women and ethnic/racial minority (ERM, which includes those who identify with any race or ethnicity other than white or Asian) students are severely underrepresented in physics courses, majors, and careers (AIP Statistics; TEAM-UP Report, 2020). The societal stereotypes and biases about who belongs in physics and can excel in it, the lack of role models, and the chilly, unsupportive, and competitive climate in physics make the playing field uneven for the traditionally underrepresented groups.

In physics education research, equity and inclusion have been studied with a focus on different traditionally underrepresented demographic student groups, e.g., women, ERM students, students with disabilities and LGBTQ+ students etc. at different points in their physics learning, e.g., high school, different levels in college including graduate school (Barthelemy *et al.*, 2016), etc. In this chapter, we discuss a theoretical framework and then review research on assessing and improving equity in physics learning environments focusing only on women and ERM *undergraduate students* with data collected from students in physics classes. In other words, our focus is *only* on research involving assessment and

https://doi.org/10.1063/9780735425514_002

strategies to make physics learning environments equitable and inclusive for women and ERM students in college-level undergraduate physics courses for both physics majors and non-majors. Prior studies have focused on assessing differences in the experiences and outcomes of different demographic groups in physics courses and physics majors as a whole to measure inequities in order to devise strategies to improve outcomes and level the playing field. As can be seen from Fig. 2.1, this research strand in physics education is relatively new and publications in this area primarily started in the 21st century.

In conducting and interpreting research on underrepresented student groups in physics, intersectionality is a useful framework because a single demographic characteristic, e.g., gender or ethnicity/race alone cannot fully explain the intricacies of the obstacles that students face (Cho *et al.*, 2013; Mitchell *et al.*, 2014; and Morton and Parsons, 2018). In particular, a combination of different aspects of an individual's social identity (e.g., gender and ethnicity/race) leads to unique levels of disadvantages that cannot be explained by simply adding together the effects of the individual components of their identity (Crenshaw, 1990). For example, according to the framework of intersectionality, in many STEM disciplines where the societal norm expects that students are white men, the experience of a Black woman is not a simple sum of the experiences of being a woman and being Black (Charleston *et al.*, 2014; and Morton and Parsons, 2018). In particular, some researchers have argued for the use of critical race theory and feminist standpoint theory in physics education research (Rodriguez *et al.*, 2022). Before proceeding further, we first explicate how we conceptualize equity in physics learning.

Our conceptualization of equity in physics learning includes three pillars: equitable access and opportunity to learn physics, equitable and inclusive learning environment, and equitable outcomes. Thus, by equity in physics learning, we mean that not only should all students have equitable opportunities and access to resources but they should also have an equitable and inclusive learning environment with appropriate support and mentoring so that they can engage in learning in a meaningful and enjoyable manner and the learning outcomes should be equitable. By equitable learning outcomes, we mean that students from all demographic groups (e.g., regardless of their gender identity or race/ethnicity) who have the pre-requisites to enroll in physics courses have comparable learning outcomes. This conceptualization of equitable outcomes is consistent with Rodriguez *et al.*'s equity of parity model (Rodriguez *et al.*, 2012). The physics learning outcomes include student performance in courses as well as evolution in their motivational beliefs such as physics self-efficacy and identity because regardless of performance, students' motivational beliefs can influence their short- and long-term retention in physics courses, majors, and careers. In other words, an equitable and inclusive learning environment should be student-centered so that all students are provided appropriate support and students from all demographic groups have equal sense of belonging regardless of their prior preparation as long as they have the prerequisite basic knowledge and skills. An equitable and inclusive learning environment would also ensure that students from all demographic groups enjoy learning physics and embrace challenges as learning opportunities instead of being threatened by them. Equitable learning outcomes for physics and other science, technology, engineering, and math (STEM) majors also include the ability of the physics courses to empower students from all demographic groups and make them

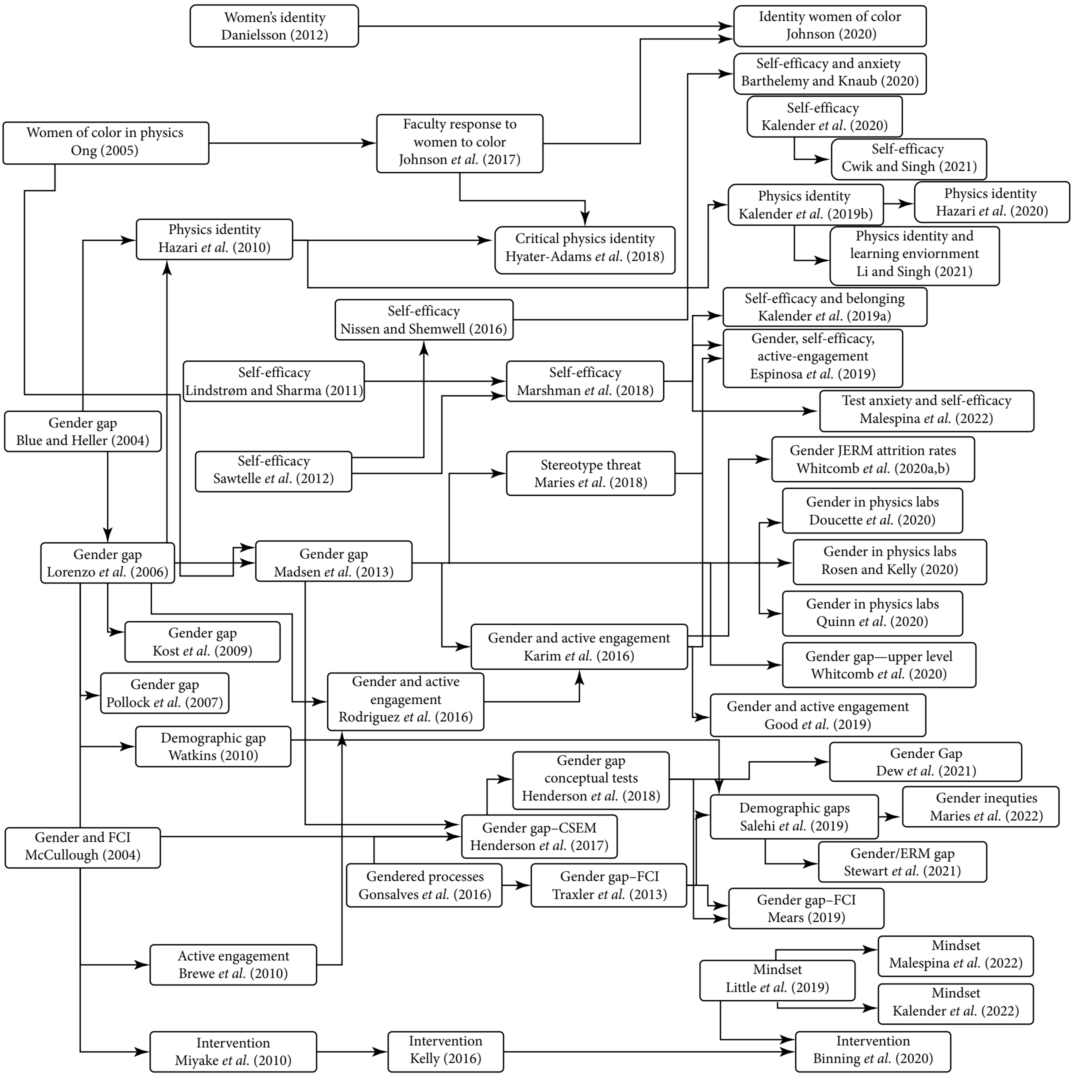


FIG. 2.1
Flowchart of physics education research on equity and inclusion focusing on women and ethnic/racial minority students with data collected from undergraduate students in physics courses. The arrows show connections between studies with some related themes.

passionate about pursuing further learning and careers in related areas. We note that equitable access and opportunity to learn physics, equitable and inclusive learning environment, and equitable outcomes are strongly entangled with each other. For example, if the physics learning environment is not equitable and inclusive, the learning outcomes are unlikely to be equitable.

Our conceptualization of equity in physics learning is mindful of the pervasive societal stereotypes and biases about physics as well as the lack of role models that can have a detrimental psychological impact on women and ERM students who are severely underrepresented. In general, when students struggle to solve challenging physics problems, they often respond in one of two ways. Some start to question whether they have what is needed to excel in physics. Others enjoy the struggle because it means that they are tackling new physics and learning. The negative reaction is a manifestation of a fixed mindset, i.e., believing that intelligence is immutable and struggling is a sign of a lack of intelligence, whereas the positive reaction is the sign of a growth mindset (the fact that your brain's capabilities can grow with deliberate effort and one can become an expert in a field by working hard). In an inequitable and non-inclusive learning environment, due to societal stereotypes and lack of role models as well as the unsupportive culture of physics, the marginalized students are more likely than others to fall prey to the fixed mindset trap and view their struggle with challenging physics problems in a negative light (Pollack, 2015). In fact, compared to most other STEM fields, the societal stereotypes are stronger in physics, a field whose history is often told through the stories of brilliant men. These stereotypes and lack of role models can also contribute to a lower sense of belonging for women and ERM students in physics learning environments unless explicit efforts are made to make them equitable and inclusive.

With our conceptualization of equity in learning, this chapter focuses on research on assessing and improving equity and inclusion using the Holistic Ecosystem for Learning Physics in an Inclusive and Equitable Environment (HELPIEE) framework adapted from the Strategies for Engaged Learning Framework (SELF) that emphasizes a holistic approach to helping all students learn physics (Marshman *et al.*, 2018b). The HELPIEE framework (see Fig. 2.2) adapts the SELF-framework to explicitly indicate that an important characteristic of an inclusive and equitable learning environment is that it will lead to equitable outcomes which are comparable for all demographic groups with regard to students' physics knowledge structure and problem-solving skills as well as physics motivational beliefs (e.g., sense of belonging, self-efficacy, identity, etc.) (Cwik, 2022d). In this framework, instructors are tasked to make their course student centered (with a special focus on marginalized students) and increase students' offensive skills, which include problem solving and knowledge, as well as offensive skills, which include students' motivational beliefs and mindset. Thus, physics courses should have learning outcomes not only based upon physics-related knowledge and skills we want students to learn but also those that focus on whether all students (and especially those from underrepresented groups) have a high sense of belonging, self-efficacy, growth mindset, and identity as people who can excel in physics.

Drawing an analogy with sports, we note that to help players excel in any game, such as tennis, coaches must ensure both good defense and offense. Likewise, helping students learn physics well requires that

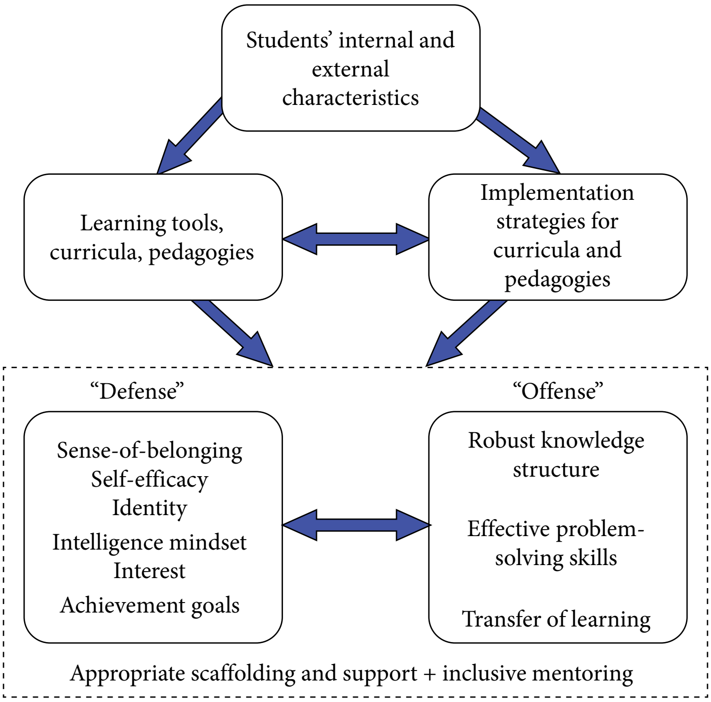


FIG. 2.2
Holistic Ecosystem for Learning Physics in an Inclusive and Equitable Environment (HELPIEE) framework.

instructors center student experiences in their holistic instructional design and equip all students with both defensive as well as offensive strategies. Instructors can strengthen students' defenses by creating equitable and inclusive learning environments in which all students have a high sense of belonging, promoting and emphasizing a growth mindset, and ensuring that all students have a high physics self-efficacy and identity. Only if instructors help develop strong defenses in students pertaining to physics learning can students effectively engage with the offense by tackling challenging problems and developing physics problem solving, reasoning, and meta-cognitive skills. If instructors do not help students develop strong defenses, many students are unlikely to risk struggling with challenging physics problems.

In the absence of physics instructors recognizing their role in centering students (which is particularly important for students from marginalized groups) and helping them build strong defenses, tackling challenging physics problems can make a student have the following type of worry: "I am struggling because I do not have what it takes to do well in physics. What is the point of even trying?" These kinds of negative thoughts can lead to a lack of engagement, even with research-based approaches to learning physics and can increase students' anxiety. This, in turn, can start a detrimental feedback loop in which negative thoughts about struggling can lead to increased anxiety, procrastination, disengagement from effective learning approaches, and failure to take advantage of resources for learning. Moreover, anxiety can rob students of cognitive resources both while learning and taking tests. This can lead to deteriorated performance, which can then lead to further negative thoughts and anxiety. Having been bombarded by societal stereotypes and biases from a young age, women and ERM students are less likely than the students from the dominant group to have strong defenses when they enter physics classes. Therefore, if the instructor does not make a concerted effort to bolster student defenses and inoculate students against stereotype threats (Dasgupta, 2011) (i.e., fear of confirming a negative stereotype about one's group), the inequitable and non-inclusive learning environment is more likely to hurt women and ERM students. On the other hand, creating an equitable and inclusive learning environment and inculcating a growth mindset, i.e., intelligence is not immutable and one can excel in physics by working hard and working smart, can go a long way in helping all students engage effectively and benefit from research-validated tools and approaches (Aguilar *et al.*, 2014; and Binning *et al.*, 2020).

Thus, consistent with our HELPIEE framework in Fig. 2.2, instructors can empower students and strengthen their defenses by creating an inclusive and equitable learning environment. With proper coaching, all students can have a high sense of belonging, be unafraid to struggle and fail, and recognize that failures are normal and should be embraced since they are stepping-stones to learning (Aguilar *et al.*, 2014, and Binning *et al.*, 2020). Although physics instructors have traditionally not considered it to be their responsibility to serve as coaches for their students (who focus on strengthening both their defensive and offensive skills), these issues involving centering of traditionally marginalized students in the instructional design are central to equity and inclusion in physics. Moreover, short classroom activities that take less than a class period at the beginning of the course can go a long way in improving students' sense of belonging and confidence, particularly for marginalized students who need the most boost due to pervasive inequities and biases (Aguilar *et al.*, 2014; and Binning *et al.*, 2020). To meet these objectives, it is critical to ensure that the learning environments in physics courses are equitable and inclusive. To ensure these equitable outcomes, physics instructors need to be trained in research-validated approaches to making the learning environments equitable and inclusive as well as in inclusive mentoring approaches. Physics instructors must also be given the opportunity to reflect upon and internalize the fact that students' physics interest and achievement goals are not fixed and can grow if the physics learning environments increase students' sense of belonging, self-efficacy, growth mindset, and physics identity in addition to helping them develop a good grasp of physics concepts.

Therefore, as shown in Fig. 2.2, the HELPIEE framework emphasizes that equitable and inclusive learning entails that students' internal and external characteristics should be at the core of making decisions about what learning tools, curricula, and pedagogies would be effective, and equally importantly, how they should be implemented and what kinds of support students should be provided. Student internal characteristics include their prior preparation and physics beliefs (which are shaped by differential prior support from societal stakeholders and differential opportunities due to lack of resources in addition to biases and stereotypes, and lack of role models) and external characteristics include support of family members or other mentors/advisors and ability to balance the course work and outside work (e.g., many students must work to support themselves), among others.

While the majority of research studies reviewed in this chapter focus on the assessment of inequities based upon gaps in performance or motivational outcomes in undergraduate physics courses and the physics major as a whole using quantitative measures (e.g., how women and ERM students may be disadvantaged if the learning environment is not equitable or inclusive), some research studies discussed focus on qualitative investigations involving ethnographic class observations and individual interviews with students and instructors. We also review research on course-level interventions for making the physics learning environment equitable and inclusive and improving the experiences and achievements of the underrepresented groups (Brewe *et al.*, 2010; Miyake *et al.*, 2010; and Binning *et al.*, 2020).

The rest of the chapter is organized as follows. We first focus on research on equity issues in quantitative research at the level of physics majors to get a big picture view of equity in undergraduate learning for physics majors. Then, we focus on quantitative research in physics courses, comparing the performance of the underrepresented groups and dominant group, followed by quantitative research on student motivational beliefs. In these contexts, we summarize findings from three types of courses: traditionally taught physics courses, research-based active-engagement courses with no special consideration for equity issues, and courses in which creating an equitable and inclusive learning environment is intentionally planned and incorporated into the course design. This is followed by research on stereotype threat since students, especially those from marginalized groups in physics, can be harmed by it. Then, we focus on qualitative research that helps us zoom in and obtain a deeper understanding of the experiences of women and ERM students and can be helpful in contemplating strategies to make physics learning environments equitable and inclusive. We close with a discussion of future directions and ongoing dialogues in the field.

2.2 QUANTITATIVE RESEARCH ON PHYSICS MAJOR-LEVEL EQUITY IN OUTCOME

In accordance with the HELPIEE framework, the research in this section focuses on measures of equity in outcomes for physics majors from different demographic groups based on their performance and grade point average (GPA). Research shows that women drop out of the physics major (and other STEM majors) with a significantly higher overall GPA than men and ERM students drop out of the physics major at a significantly higher rate than white students (65% of ERM students vs 34% of white students) (Whitcomb and Singh, 2021; and Maries *et al.*, 2022). Physics has one of the lowest enrollment as majors of all STEM disciplines, and student attrition after declaring a major is the highest from physics compared to other disciplines (Whitcomb and Singh, 2021; and Maries *et al.*, 2022). Research also shows that physics is a STEM discipline that does not attract a second wave of majors in the third year and beyond, i.e., those who initially declare other STEM majors do not switch to physics in later years unlike most other STEM majors, which have a bi-directional flow of students in later years (Whitcomb and Singh, 2021). Moreover, it is important to keep in mind that this research on a documented unidirectional flow of physics majors out of physics (Whitcomb and Singh, 2021; and Maries *et al.*, 2022) does not include the attrition of students who were intending to major in physics in their first year (e.g., due to their better physics experiences in high school physics) but changed their mind after taking their college introductory physics courses even *before* declaring the physics major (since many students in U.S. colleges declare the physics major at the end of their first year or in their second year).

In addition, students' experiences in introductory courses may affect their decisions in upper-level courses (Rodriguez *et al.*, 2016; and Whitcomb and Singh, 2020). In one study on introductory and advanced physics and math courses, gender differences in physics course performance (controlling for high school GPA) were found only in introductory physics courses (Whitcomb and Singh, 2020). Additionally, the introductory courses did not predict performance in advanced physics courses. Therefore, women could be making decisions about whether physics and related disciplines are the right fields to major for them based on the gender performance gap, e.g., in the introductory physics courses. In another study, researchers investigated the student performance and persistence in upper level physics courses for students who experienced active learning (specifically, modeling instruction and the investigative science learning environment) in introductory physics courses (Rodriguez *et al.*, 2016). In this study, women were more likely than men to graduate with a physics degree and they were just as likely as men to pass the upper-level courses. The highest risk of failure for women and men was in the first semester of upper division courses.

Another study investigated grades earned by students categorized by some of their demographic factors including gender, ERM student status, low-income status, and first-generation college student status. ERM students experienced the largest penalty in their STEM GPA and overall GPA compared to non-ERM students (Whitcomb *et al.*, 2021). While women had a higher overall GPA than men, the gender gap either reduced or disappeared for students' STEM GPA depending on the demographic group (ERM student status, low-income status, and first-generation college student status). In addition, when investigating students' STEM GPA for physical science (chemistry, computer science, engineering, mathematics, and physics) majors, ERM students experienced the largest penalty in their STEM GPA and, in general, students who were from intersecting marginalized demographic groups (for example, low income and ERM) had larger penalties than students with the most advantages or students with another single demographic group disadvantage.

2.3 QUANTITATIVE RESEARCH ON COURSE-LEVEL EQUITY IN OUTCOME: PERFORMANCE

In accordance with the HELPIEE framework, the research in this section focuses on measures of equity in outcomes for students from different demographic groups (not necessarily physics majors) measured by their performance in physics courses. Early research on equity and inclusion in physics focused on documenting and/or reducing the gender gap in performance in physics courses. Some of the earlier work focused on the gender gaps in concept inventories such as the Force Concept Inventory (FCI) and the Conceptual Survey of Electricity and Magnetism (CSEM) (Blue and Heller, 2004; McCullough, 2004; Lorenzo *et al.*, 2006; Coletta *et al.*, 2012; Madsen *et al.*, 2013; Henderson *et al.*, 2017; Traxler *et al.*, 2018;

and Mears, 2019). Some of the research has shown that when high school backgrounds and pretest scores were matched, there was no difference in post-test physics performance (Blue and Heller, 2004). Other research studies have found that particular questions on the FCI or CSEM may be biased against women (McCullough, 2004; Henderson *et al.*, 2017; Henderson *et al.*, 2018; and Traxler *et al.*, 2018). In addition, one study investigated gender differences in midterm and final exam scores and whether those gender differences were correlated with final course grades for introductory physics courses (Dew *et al.*, 2021). Results show that performance on exams and final grades was weakly dependent on student gender (Dew *et al.*, 2021). Research suggests that gender differences in performance may be caused by societal stereotypes and biases about physics (Blue and Heller, 2004). The gender differences in scores are most likely due to a combination of many factors rather than one factor that can be easily modified (Madsen *et al.*, 2013). One study found that high school factors (including SAT scores, enrollment in calculus courses, and high school grades) predict student performance in introductory physics courses and that the pedagogy used in high school physics courses could differentially predict male and female students' performance (Hazari *et al.*, 2007).

A growing body of research has documented the gender gap when active learning pedagogies are implemented in the physics classroom. Lorenzo et al and Coletta *et al.* found that interactive engagement pedagogy reduces the gender gap in performance in the physics course and on the FCI, respectively, and both women and men benefit from interactive engagement pedagogy (Lorenzo *et al.*, 2006; and Coletta *et al.*, 2012). However, other studies have found that interactive engagement pedagogies do not reduce the gender gap and may worsen the gap (Pollock *et al.*, 2007; Kost *et al.*, 2009; Brewe *et al.*, 2010; Karim *et al.*, 2018; and Good *et al.*, 2019). That is, active engagement in an inequitable and non-inclusive learning environment can lead to inequitable outcomes, e.g., it can lead to an increased gender performance gap. For example, one study investigated the gender gap in an introductory physics 1 course on the FCI when modeling instruction (MI) was implemented as opposed to lecture-based courses (Brewe *et al.*, 2010). Results show that while students in the MI course outperformed students in the lecture-based course, the gender gap in women's scores on the FCI increased at the end of the course. Another study investigated students' conceptual test performance in multiple introductory courses in which either partially or fully interactive classroom techniques were used (including student discussions on ConceptTests and Tutorials, among others) (Pollock *et al.*, 2007). This study showed that there was a learning gap between male and female students in these courses and that in some instances, the gap increased at the end of the course. Additionally, in another study (Maries *et al.*, 2020), the gender gap on the CSEM survey in calculus-based college introductory courses that primarily make use of lecture-based (LB) instruction (i.e., instructor lectured 90% or more of the class) was compared with courses that make significant use of evidence-based active-engagement (EBAE) strategies. The EBAE courses included flipped courses in which students watched lecture videos at home and answered a pre-lecture assignment before coming to class, and class time was used for clicker questions involving peer discussions and lecture demonstrations preceded by questions and collaborative problem solving in which students worked in groups of two to three on quantitative problems. Research shows that the gender gap on the CSEM remains relatively constant in LB courses (4% at the beginning and 6% at the end), and both male and female students exhibit similar normalized gains on the CSEM (18% and 22% for female and male students, respectively). For EBAE courses, the gender gap increases significantly from 4% to 10%, which is also reflected in the effect sizes comparing male and female student CSEM performance: 0.27 on the pretest and 0.54 on the posttest, i.e., the effect size doubles. Also, the data suggest that while both men and women benefit from EBAE instruction (larger normalized gains in EBAE courses compared to LB courses), male students benefit disproportionately more than female students: the normalized gain for male students in EBAE courses is 39% compared to only 28% for female students. Here, the normalized gain is defined as (post%−pre%)/(100%−pre%).

One study suggests that the gender gap in interactive physics courses may be due to differences in prior physics and math knowledge as well as differences in attitudes and beliefs about physics (Kost *et al.*, 2009). However, in accordance with our HELPIEE framework, the instructors must take responsibility and strive to make the physics learning environments equitable and inclusive so that all students, regardless of their prior knowledge and beliefs (which are often the result of stereotypes, biases, and differential opportunities) can excel in physics courses.

While there have been many studies focused on gender issues in physics, fewer studies have focused on race/ethnicity in physics. Similar to the gender gap in performance in some studies, ERM students tend to have lower performance in introductory physics courses overall and on conceptual tests than their white peers (Brewe *et al.*, 2010; Watkins, 2010; Salehi *et al.*, 2019; and Van Dusen and Nissen, 2020). This is the case even in classes that use active engagement. In one study, ethnic and racial minority students entered the introductory physics courses with lower conceptual understanding, and the gap was maintained until the end of the introductory physics course sequence when active learning (specifically modeling instruction) was implemented (Brewe *et al.*, 2010). Other studies show that part of the gap may be explained by differences in the prior preparation. For instance, one study showed that the gaps in ethnic and racial minority students' final exam scores disappeared when students' ACT/SAT math scores and pre-test scores on a conceptual test were controlled for (Salehi *et al.*, 2019). In addition, in another study, ERM students experienced a larger penalty in their STEM GPA than non-ERM students, and ERM students with additional disadvantages due to socioeconomic status or first-generation college status were further penalized in their average GPA (Whitcomb *et al.*, 2021). Stewart *et al.* investigated physics performance differences in relation to gender, ERM status, and status as first-generation college students (FGCS) (Stewart *et al.*, 2021). Results showed significant differences in gender, ERM status, and FGCS status on the pre- and post-conceptual test and significant differences were found in students' course grades for ERM students and those with FGCS status (Stewart *et al.*, 2021). In addition, path analysis was used to examine the relationship between demographic factors, academic factors (SAT math scores, and pre-conceptual test scores), and course performance (post-conceptual tests and course grades). For ERM and FGCS students, differences in their pre-conceptual test scores and ACT math scores explained most of the difference in course achievement scores (Stewart *et al.*, 2021).

2.4 QUANTITATIVE RESEARCH ON COURSE-LEVEL EQUITY IN OUTCOME: MOTIVATIONAL BELIEFS

In accordance with the HELPIEE framework, the research in this section focuses on measures of equity in outcomes for students from different demographic groups measured by their motivational beliefs. Starting around 2010, researchers began to investigate the motivational beliefs of students, including their physics identity and self-efficacy. Early work investigated gender gaps in these motivational constructs, especially in students' self-efficacy (Lindstrøm and Sharma, 2011; and Sawtelle *et al.*, 2012). Self-efficacy is one's belief that they can succeed in a particular activity or course (Bandura, 1977, 1994). Self-efficacy is an important motivational construct since it is one of the primary dimensions of physics identity and it influences students' engagement, learning, and persistence in science courses (Lindstrøm and Sharma, 2011; and Kalender *et al.*, 2020). Sawtelle *et al.* investigated sources of students' physics self-efficacy by gender and found subtle distinctions in the predictive ability of the sources of self-efficacy in women and men (Sawtelle *et al.*, 2012). According to Badura's social cognitive theory, self-efficacy may be derived from four sources: mastery experiences, vicarious learning experiences, social persuasion experiences, and physiological state (Bandura, 1994). Mastery experiences are important because successful completion of a task should have a strong positive influence on one's confidence to complete a similar task. Vicarious learning experiences occur when observing someone else's success on a task influences their own belief in their ability to perform a similar task. Social persuasion experiences show that verbal suggestions from others, such as words of encouragement, can result in an increase in one's self-efficacy. Lastly, one's physiological state can act as a mediating source to amplify or undermine one's confidence in one's ability. Sawtelle *et al.* found that the probability of passing an introductory physics course for women relies primarily on the vicarious learning experience source, while it relies on mastery experiences for men (Sawtelle *et al.*, 2012).

In general, self-efficacy of both men and women decreases throughout introductory physics courses and self-efficacy decreases more for women than men (Marshman *et al.*, 2018a). This could be detrimental to women since self-efficacy and test anxiety may be related to each other (Malespina and Singh, 2022). In one study, women had lower self-efficacy and higher test anxiety than men, and self-efficacy mediated the relationship between test anxiety and high-stakes assessment test scores (Malespina and Singh, 2022). The gender gap in self-efficacy is found in both traditional and most interactive engagement courses (Nissen and Shemwell, 2016). However, one study found that in a class that employed team- and project-based physics learning, the gender gap in self-efficacy disappeared at the end of a physics class (Espinosa *et al.*, 2019). In general, this gender gap in self-efficacy persists for men and women even when controlling for students' performance in introductory calculus-based physics courses (Marshman *et al.*, 2018a). Specifically, women who received A's in the course had the same self-efficacy as men who received C's in the course (Marshman *et al.*, 2018a). For engineering students, the gender gap in physics self-efficacy does not close by their fourth year, whereas it closes in other STEM subjects (Whitcomb *et al.*, 2020). Self-efficacy may predict students' future careers in physics as well (Barthelemy and Knaub, 2020). In particular, in a study of Finnish undergraduate students, women had lower self-efficacy and higher anxiety about physics than male students, and their self-efficacy predicted student goals of going to graduate school (Barthelemy and Knaub, 2020).

Around the same time physics self-efficacy was investigated, researchers also explored students' physics identity. Physics identity is defined as identifying with physics, i.e., whether students see themselves as physics people (Hazari *et al.*, 2010; and Monsalve *et al.*, 2016) or those who can excel in physics. In 2010, Hazari *et al.* took advantage of Carlone and Johnson's framework for science identity (Carlone and Johnson, 2007) to formulate a framework of physics identity and began to investigate high school students' physics identity and college students' engineering identity (Hazari *et al.*, 2010). In Hazari's framework, "competence" and "performance" were defined as students' beliefs in their ability to understand the subject and students' belief in their ability to perform physics tasks. Additionally, recognition was framed as recognition by others as being a good physics student. Lastly, a fourth dimension, interest, was added to the framework since students can have highly varying levels of interest in physics (Hazari *et al.*, 2017; and Hazari and Cass, 2018). In more recent studies by Hazari *et al.* of introductory students, performance and competence were combined into one variable (Hazari *et al.*, 2020). In a slightly reframed version of Hazari *et al.*'s physics identity framework by Kalender *et al.* (2019b), performance/competence was framed as self-efficacy, which is closely related to competency belief and recognition was renamed "perceived recognition" for clarity to investigate introductory students' physics identity (Kalender *et al.*, 2019b). Hyater-Adams *et al.* combined the frameworks of physics identity and racialized identity to create a Critical Physics Identity framework to understand the experiences of Black students in physics (Hyater-Adams *et al.*, 2018).

Our prior individual interviews suggest that students' perceived recognition by instructors and teaching assistants (TAs) impacts their self-efficacy and interest in physics, e.g., see Doucette *et al.* (2020) and Doucette and Singh (2020). In addition, in order for students to feel validated and recognized by their instructors, instructors need to both explicitly recognize students (for example, by verbally acknowledging the progress and success of their students) and implicitly recognize students (for example, by setting high standards for all of their students and making it clear to students that they all have what it takes to excel if they work hard and work smart and take advantage of all of the resources) (Wang and Hazari, 2018). However, studies have shown that without intentional strategies to create an equitable and inclusive learning environment, female students do not feel recognized appropriately even before they enter college (Archer *et al.*, 2017; and Kalender *et al.*, 2019a), which is at least partly due to the societal stereotypes and biases about who belongs in physics and who can excel in it that women are bombarded with over their lifetime.

Students' sense of belonging in physics courses has not been studied as extensively. However, it has been shown that students' sense of belonging in calculus-based introductory physics courses for physical science and engineering majors is so closely tied to their self-efficacy that it was difficult to separate them as distinct factors in factor analysis (Kalender *et al.*, 2019a). In addition, it is shown to be a predictor of students' physics identity for senior physics majors (Hazari *et al.*, 2020), and a predictor of students' grades in an introductory physics 1 course for students on the bioscience track (Cwik and Singh, 2022c). Another study showed that female participants in a Physics Olympiad competition who endorsed negative stereotypes about female talent for physics felt a lower sense of belonging in physics (Ladewig *et al.*, 2020). One study investigated factors that affect women's sense of belonging in physics, including the lack of role models in physics and stereotypes about who can succeed in physics (Lewis *et al.*, 2016). Some suggestions for instructors to improve women's sense of belonging in physics include sending messages that concerns about belonging are normal and fade with time, identifying and tempering cues that perpetuate the "geeky" scientist stereotype, and openly endorsing effort and hard work over brilliance (Lewis *et al.*, 2016).

Additionally, although general mindset research has been ongoing for many decades (Dweck, 2008), mindset applied to college physics has only recently started to become more widely studied in the physics discipline (Kalender *et al.*, 2022; and Malespina *et al.*, 2022). One study in an introductory calculus-based physics course laid out a framework in which there are four distinct mindset views (i.e., whether students' ability is fixed or malleable, whether their intelligence can grow from effort, and similarly, whether that view is about themselves or for the general population) and investigated how they predict physics course grade (Kalender *et al.*, 2022). The researchers found that women were more likely to believe that innate talent was needed for them to excel in physics and that they might not be gifted. In addition, mindset was a stronger predictor of physics grade of students' malleable mindset views about themselves ("my ability") than the other mindset groups. In another study in an introductory calculus-based physics course, there were only gender differences in students' "my ability" mindset group at the beginning of the course (Malespina *et al.*, 2022). However, gender differences developed in each mindset view at the end of the course, and the gender difference in the "my ability" category increased over the course. In addition, "my ability" was the only mindset factor that predicted course grades.

While most of the studies on motivational beliefs have been conducted in calculus-based introductory physics classes in which women are severely underrepresented, similar findings have been reported in algebra-based physics courses for students interested in health professions in which women are not underrepresented. Although women are not underrepresented in these physics courses, societal stereotypes and biases internalized by female students over their lifetime can still impact their motivational beliefs about physics. In studies of introductory physics courses for bioscience majors, women had lower motivational beliefs than men. Although women outnumber men in these introductory physics courses, women had lower self-efficacy than men controlling for the grade they received in the physics courses. In particular, women who received an A in physics 1 and 2 courses (defined as A+, A, or A−) had similar self-efficacy to men who received a B (defined as B+, B, or B−) (Cwik and Singh, 2021a). This trend is similar to the trends in calculus-based introductory physics courses in which women with A grades have the same self-efficacy as men with C grades (Marshman *et al.*, 2018a). However, instructors may have the ability to help improve students' motivational beliefs in the introductory courses. In studies conducted in calculus-based introductory physics courses and introductory physics courses for bioscience majors, students' perception of the inclusiveness of the learning environment factors (consisting of recognition by others including physics course instructors and teaching assistants, interaction with their peers, and sense of belonging) predicted gender differences in students' motivational outcomes at the end of the physics course such as self-efficacy, interest, and physics identity (Cwik *et al.*, 2020; and Li and Singh, 2021, 2022).

In addition, some research studies have focused on students' motivational beliefs in the lab context. One study investigated students' two different group work styles in introductory lab courses: Group A in which each student takes on a different task but spends equal time on it, and Group B in which students divide the work equitably and each student participates in every aspect of the work (Doucette and Singh, 2022). The findings show that while students prefer Group A style work, students who participated in Group B style work were more likely to report that interacting with their peers increased their physics interest and women were more likely to report that peer interactions increased their self-efficacy (Doucette and Singh, 2022). In addition, in another study in the lab context, an online, hands-on laboratory option was implemented to investigate gender differences in students' epistemological beliefs, socialization, and help-seeking in the laboratory (Rosen and Kelly, 2020). Results show that men had higher epistemological beliefs, women reported a greater willingness to seek assistance from instructors and peers, and there was no difference by gender in socialization. The students in the in-person labs placed a higher value on socialization in the lab, while there was no difference in epistemological beliefs or help-seeking. However, when comparing male and female students in each lab type, there were no significant gender differences in the three factors.

Several factors have been proposed to explain the gender difference in students' motivational beliefs (including self-efficacy and identity) and on conceptual tests. Some studies suggest that differences in the prior preparation for various reasons may account for the gender difference (Kost *et al.*, 2009; and Kost-Smith *et al.*, 2010). In addition, societal stereotypes and biases who belongs in physics can negatively impact women in physics courses (Blue *et al.*, 2018). Some of the elements of the environment in many science classrooms that can negatively impact women include a lack of female role models, pedagogy that favors male students' interests, and a "chilly climate" for women (Blickenstaff, 2005).

In addition, the TEAM-UP report lists five factors essential to improving African American students' persistence and success in physics, including their belonging, physics identity, academic support, personal support, and leadership and structures (TEAM-UP Report, 2020). In order to improve African American students' sense of belonging, the report includes suggestions such as establishing clear rules in common spaces to ensure everyone is welcome, assisting students in finding the support they need inside and outside the department, and consistently communicating norms and values of respect and inclusion. To improve physics identity, suggestions to physics departments include diversifying their faculty with respect to race/ethnicity/gender, emphasizing the ways a physics degree empowers graduates to improve society, and discussing a broad range of career options with undergraduate students. In addition, to increase the other three factors, the report suggests that departments develop evidence-based actionable plans to increase the persistence of all students to physics degrees, help students take advantage of campus resources for funding such as conference travel, and set norms of inclusion and belonging in the department.

2.5 STEREOTYPE THREAT

In accordance with the HELPIEE framework, the research in this section focuses on how the stereotype threat can be particularly detrimental to marginalized students if the physics learning environments are not equitable and inclusive. The stereotype threat is the anxiety associated with confirming a negative stereotype (e.g., related to women and ERM students in physics) resulting in reduced performance for the stereotyped group (Steele and Aronson, 1995). Prior studies in the context of mathematics have found that activation of a negative stereotype about a group or stereotype threat, e.g., asking test takers to indicate their ethnicity before taking a test, can lead to a deteriorated performance of the stereotyped group. For example, in the high school physics context, Marchand and Taasoobshirazi (2013) conducted research that suggests that a stereotype threat is automatically triggered in a physics test-taking situation due to prevalent societal stereotypes. They used three different manipulations immediately before students took a four question quantitative physics test: (i) an explicit, (ii) an implicit, and (iii) a nullified stereotype threat condition in which students were either told that (i) female students had performed worse than male students on this test, (ii) not told anything, or (iii) told that the test had been found to be gender neutral. While male students performed similarly in all three conditions, female students in the explicit and implicit stereotype threat conditions had comparable performances and performed statistically significantly worse than female students in the nullified condition. The researchers interpreted this result to suggest that a stereotype threat is automatically triggered in a test-taking situation for women in physics courses.

In the context of college physics courses, the effect of stereotype threat has been studied to understand its impact on student performance, e.g., on standardized tests. In one study (Maries *et al.*, 2018), the pretest and post-test performance of female and male students on the Conceptual Survey of Electricity and Magnetism (CSEM) was analyzed in an introductory algebra-based course in these two conditions: students were or were not asked to provide gender information before taking the CSEM gender salient or not salient condition, respectively. There were no statistically significant differences between the performance of male or female students under the two conditions (e.g., female students who wrote their gender before taking the CSEM did not perform worse than female students who wrote their gender after taking the CSEM) in the pretest or the post-test. One potential explanation for this is that the stereotype threat may be activated without needing to ask about gender, similar to the Marchand and Taasoobshirazi study (Marchand and Taasoobshirazi, 2013).

In a related study (Maries *et al.*, 2018), investigators focused on stereotype threat associated with gender stereotypes in physics and its impact on student performance on the CSEM in calculus-based college introductory courses taken by engineering, physical science, and mathematics majors in which female students are severely underrepresented. In particular, the study investigated the extent to which agreeing with a gender stereotype (i.e., I expect men to generally perform better in physics than women) correlates with performance on the CSEM. The demographic information of students in these courses was gender—67% male and 33% female, race—77% White, 11% Asian, 4% Latinx, 4% Multiracial, 3% Black, and 1% Other. It was found that women who agreed with the gender stereotype performed statistically significantly worse than women who did not agree with the gender stereotype at the end of a year-long calculus-based physics course sequence, even though there was no difference in their performance in the pretest given at the beginning of the course (Maries *et al.*, 2018). Cognitive science suggests that the anxiety associated with conforming to a stereotype is essentially a threat, and it can take up part of the working memory, thus robbing an individual of cognitive resources that could be used for problem solving and learning. Moreover, the anxiety can also lead to procrastination or less time spent on learning as well as reduced engagement and use of effective study strategies and asking for help. As noted, prior research has suggested that a certain level of stereotype threat may be implicitly present for female students in an introductory physics course. These findings suggest that female students endorsing a gender stereotype may be undergoing additional stereotype threats over and above what might already be present for many women in physics courses. As suggested by prior research (Maries *et al.*, 2018), over the course of the semester, this can have a significant negative impact, especially in a calculus-based introductory physics course in which women are underrepresented.

2.6 QUALITATIVE METHOD RESEARCH

In accordance with the HELPIEE framework, the qualitative research in this section focuses on obtaining a deeper understanding of the physics learning environments and the physics culture in order to create an equitable and inclusive learning environment and transform the physics culture so that it centers on the experiences of marginalized students. In particular, qualitative research can be a powerful tool for unpacking the mechanisms underlying quantitative research discussed in the preceding sections and understanding the experiences of students (especially women and ERM students we focus on here). Thus, prior research studies using qualitative methods, e.g., interviews with women and ERM students we discuss here, focus on understanding their experiences in physics classes in order to improve the learning environments. For example, qualitative research has focused on students' mindset and how mindset beliefs studied in the literature in other disciplines are not always consistent with challenges in college physics courses (Little *et al.*, 2019).

A series of qualitative research studies conducted in the past decade have focused on the experiences of women in physics. For example, one study focused on the masculine nature of doing physics in a variety of contexts (Gonsalves *et al.*, 2016) and how it negatively impacts women. Another qualitative research involved interviews with five women physics students to understand their identity as physicists (Danielsson, 2012). The study showed that some of the women adapted themselves to relate to the masculine norms of the discipline. For example, one of the interviewed women mentioned that she was likely to tinker with lab equipment and referred to herself as "laddish," which made it easier for her than other women to fit within the boundaries of the existing physics culture. Some of the women also reflected upon the norms and expectations about how women are supposed to behave in a physics learning space, e.g., how they were more likely to take on the secretarial role in physics labs since their male partners wanted to be the tinkerers. In addition, research involving interviews and ethnographic observations to understand gendered task division shows that women often have to position themselves as secretaries or project managers in the labs and do gender and physics simultaneously (Doucette *et al.*, 2020; and Quinn *et al.*, 2020).

Qualitative research focusing on race and ethnicity in the context of physics learning environments has often utilized a critical race theory lens to interview women of color to understand their experiences through the physics curriculum (Ong, 2005; Rosa and Mensah, 2016; and Quichocho *et al.*, 2019). For example, in one study, six Black women were interviewed to understand the obstacles faced in their career paths from the beginning all the way to when they were graduate students, such as socialization in STEM (Rosa and Mensah, 2016). Many of the women had similar experiences in that they were influenced to major in physics due to being exposed to a science environment at a young age from after-school or summer school programs and were able to conduct physics research over the summer in their undergraduate studies. However, most of the women experienced isolation in their graduate studies, particularly in study groups. Strategies used to overcome the obstacles faced in choosing physics over other STEM fields and in their career included afterschool activities that focus on scientific practices in high school, college recruitment specifically targeting underrepresented groups, financial aid, and creating an inclusive and supporting environment (Rosa and Mensah, 2016). In another study, ten women of color were interviewed to understand how their sense of belonging and competence is questioned due to the intersection in three realms: their field of study (physics), gender, and race/ethnicity (Ong, 2005). It was concluded that in order to retain more women of color in the field, physics departments should focus on reforming hiring and recruiting policies to recruit more women and ERM students, structurally and financially supporting their membership in organizations (such as the formal groups for women and ERM students), and creating a more hospitable environment for all students by improving the pedagogical, social and cultural practices that attract potential science majors (Ong, 2005).

In another study, women of color and lesbian, gay bisexual, transgender, or queer women were interviewed about their identity as a physicist (Quichocho *et al.*, 2019). When asked to describe a physicist, the women mentioned common stereotypes about who can do physics (such as being a white, male, genius to succeed in physics) that have prevented them from identifying as a physicist and questioned the necessity of a formal degree to be a physicist. However, most of the women were able to eventually reject the common stereotype and self-identify as a physicist and thus came to accept that others saw them as a physicist as well. The study highlighted the importance of personally identifying as a physicist for empowerment and belonging and the importance of recognition by others as a component of identity (Quichocho *et al.*, 2019). While most of this type of research has focused on students' experiences while they are still undergraduates, some of the research has focused on graduate students in which they were asked to reflect upon the difficulties ERM students face in contemplating applying for graduate degrees in physics. For example, one study investigated student responses on an application question to be a part of the APS Bridge program and interviewed 9 students who were accepted into the program to understand the barriers that ERM students face when applying to graduate school (Cochran *et al.*, 2018). Findings showed that some of the barriers that ERM students face include the Graduate Record Exams (GRE), student research experience, student grades/GPA, deadlines for applying to physics graduate programs, and financial concerns (Cochran *et al.*, 2018).

In order to support equity and inclusion for intersectional identity in academic settings, such as in physics classrooms, the Intersectionally Conscious Collaboration (ICC) protocol was created (Boveda and Weinberg, 2020). The ICC protocol includes six elements that allow physics educators to locate and address biases in pedagogical practices and better design learning experiences that engage and motivate all students.

There are very few examples of exemplary physics programs. One example was highlighted in a qualitative study that investigated a physics department in which women and women of color feel successful and have a high sense of belonging to understand their physics identity in that setting (Johnson, 2020). The study consisted of interviews with students and faculty members and the researcher attending physics classes as observers. Important components that promoted these women of color's physics identity included students working collaboratively together, physics faculty members taking responsibility for group work to go smoothly, protecting students from racist and sexist microaggressions, and all stakeholders (faculty members and students) believing that success in physics is a result of hard work instead of innate intelligence (Johnson, 2020).

2.7 INTERVENTIONS

In accordance with the HELPIEE framework, the research in this section focuses on interventions that are designed to make the course outcomes equitable. In particular, in addition to research into assessing equity and inclusion in undergraduate physics, there has also been research into implementing changes in the classroom to make them more equitable and inclusive. Socio-psychological classroom interventions, e.g., those focusing on self-affirmation or sense of belonging and mindset interventions, have been shown to improve the outcomes for women in physics courses (Aguilar *et al.*, 2014; and Binning *et al.*, 2020). One intervention, called values affirmation, conducted in an introductory physics course involved students writing about their most important values (such as connections with friends or family) for 15 min twice at the beginning of the course (once during the first recitation of the semester and once for online homework shortly before the first midterm) (Miyake *et al.*, 2010). Value affirmation can buffer people against psychological threats, such as the stereotype that men are better than women at math and science. At the end of the course, the gender gap in performance was reduced and benefited women who tended to endorse the stereotype that men do better than women in physics. However, in some other research studies, the values affirmation intervention did not reduce the gender gap (Gutmann and Stelzer, 2021). Therefore, more research should be conducted to determine the factors that make these types of interventions successful in a particular type of physics class for different demographic groups.

Other types of short interventions have been implemented at the beginning of the semester in introductory physics courses to normalize adversity and create an equitable and inclusive learning environment in which students from marginalized groups have a high sense of belonging and feel that it is safe to engage in collaboration and discussions with peers and instructors. With this type of short intervention, the performance gap between the underrepresented and dominant groups can sometimes be significantly reduced (Binning *et al.*, 2020). For example, a short ecological belonging/mindset intervention by Binning *et al.*, which only requires half of a single recitation class period at the beginning of the semester, eliminated the gender gap in calculus-based introductory physics performance (Binning *et al.*, 2020). In addition, non-white students performed better in the intervention condition than in the control condition. However, the intervention did not statistically eliminate the gap between white and non-white students, which may partly be due to the low numbers of non-white students.

This intervention (Binning *et al.*, 2020) was conducted in a required introductory calculus-based physics course, which is typically taken by physical science and engineering majors in their first year and their first semester in college. Two female physics graduate students were trained to facilitate the half-hour activity at the beginning of the semester in half of the recitations that were randomly selected. The facilitators introduced it as an activity that would help the physics department understand student concerns and how to foster better learning environments. At the beginning of the first recitation class in which the activity took place, students were handed a piece of paper and asked to write down their concerns about being in the physics course. Then, they were shown some quotations from both male and female students from previous years who did very well in physics but also had similar concerns. The quotes emphasized the importance of working hard and working smart, learning from one's mistakes, and taking advantage of all learning resources available to them because that is the way to perform well in physics. Then students were asked to get together in small groups to discuss what they wrote and why don't students realize that struggling is normal; during this session, they generally learned that their peers in the class had similar worries. A general class discussion followed in which the groups summarized their discussions, with explicit emphasis on the fact that adversity is common in college physics courses, but it is temporary. The facilitators re-emphasized that students should embrace challenging physics problems and use their failures as bridges to learning. Finally, using the principle of "saying is believing," students were asked to write a short letter telling a future student about strategies for excelling in their physics classes. As noted, it is heartening is that this short intervention closed the gender performance gap and greatly reduced the gap between the non-white and white students in performance compared to the group in which this short intervention did not take place. One reason these types of interventions show benefits for marginalized students is that students' sense of belonging and other motivational beliefs are strongly intertwined with feeling safe, increased cognitive engagement, reduced anxiety, and learning.

In order for these types of interventions to succeed, however, a variety of factors must be considered and carefully implemented (Aguilar *et al.*, 2014). Some of the elements that must be considered are that the intervention must deal with specific concerns students have, the message should be delivered without singling out any particular group, the intervention must use methods that psychologists have found to be long lasting, and they should not be framed explicitly as interventions but activities that are components of the course. In addition, research shows that there are multiple strategies that can improve women's participation in physics, including creating a positive learning atmosphere, providing encouragement and support to women and emphasizing the societal benefits of physics (Kelly, 2016),

Apart from interventions, another way instructors can improve the learning environment in their courses is by providing mentoring and support for underrepresented students (TEAM-UP Report, 2020). Instructors can set an equity goal for their class to explicitly track whether the demographic differences in their courses are getting better. Additionally, the mindset of the instructor in a course also plays a pivotal role in predicting student achievement. In a study of 150 STEM instructors by Canning *et al.*, courses taught by instructors with a fixed mindset had twice as large an achievement gap as courses taught by instructors with a growth mindset (Canning *et al.*, 2019). Only if the instructors themselves have a growth mindset about their students' ability can they come across as trustworthy and authentic to their students when communicating with them. Then, they can credibly and authentically emphasize a growth mindset to their students, that the physics they are learning is mastered through hard work and deliberate practice and not through innate talent or genius (Johnson *et al.*, 2017).

2.8 CRITIQUES, ONGOING DIALOGUES, AND FUTURE DIRECTIONS

One critique of most of the current work on gender in physics is that gender is put into binary categories when it is more complicated and on a spectrum. One recommendation among others is to conduct more qualitative research, based on a methodology that addresses the complexity of gender (Traxler *et al.*, 2016). Most researchers conducting this type of research now acknowledge that gender is not a binary construct, but since most of the data from students are provided by the university in a binary form, it is often used that way, particularly in quantitative research. One potential solution would be to collect the information from the students directly (for example, on a survey with multiple options for gender). However, physics education researchers must consider that they may need a specific type of IRB approval and also ask the question about gender at the end of the survey in order to reduce the stereotype threat.

Another critique is that the gender gap in performance is framed as a comparison of women with men, which could be interpreted as "women should be more like men," which is a deficit model (Traxler *et al.*, 2016). A similar critique has been put forward for the gap between white students and ERM students. The recommendation is to move beyond the "gap" framework (among others) (Traxler *et al.*, 2016). Regarding comparing women with men or ERM students with white students in quantitative comparisons in physics courses that show gaps, these kinds of studies are important for revealing inequities in the physics learning environments. In particular, the framework of the research plays a pivotal role in whether the model is a student deficit model or a course deficit model (i.e., the learning environment is not equitable and inclusive and is disadvantaging the underrepresented groups and leading to inequitable outcomes). In particular, these inequities are typically caused by differential opportunities for students based upon their privilege and societal stereotypes and biases about who can succeed in physics that can accumulate over the students' lifetime. For instance, even before stepping into undergraduate physics courses discussed here, throughout K-12 education, women and ERM students are often not treated the same way as the white male students in physical science courses by their teachers and high school counselors, and they also often give them differential advice. Even TV shows like The Big Bang Theory as well as the interactions of students in museums with adults perpetuate the stereotypes based on gender, ethnicity, and race (Crowley *et al.*, 2001). Moreover, most famous physicists are white men (including almost everyone mentioned in physics textbooks) and so women and ERM students do not have as many role models to show them that they can be successful in physics courses. It is not surprising then that in an inequitable and non-inclusive learning environment, even in introductory algebra-based physics courses for bioscience majors in which women are not underrepresented, women have lower motivational beliefs at the beginning of the course due to the societal stereotypes and biases (Cwik and Singh, 2021a, 2021b, 2022a, and 2022b). It is important to center instruction in the physics courses in ways that focus on creating an equitable and inclusive learning environment and counter the impact of these pervasive societal stereotypes and biases.

Another critique pertains to the narrow scope of the demographics in physics education research (Kanim and Cid, 2020). In particular, physics education research disproportionately focuses on students in introductory calculus-based courses and at institutions that have a smaller population of ERM students than the overall college-bound population. There is less research on high school students, students at two-year colleges, and racially diverse colleges, e.g., minority serving institutions (MSI). Additionally, most of the research has been conducted in traditionally taught courses where there are limited opportunities for active learning in the classroom. Therefore, future studies in physics education should include a wider variety of students, courses (such as physics courses with evidence-based active learning and those with explicit focus on centering student experiences), and institutions.

Additionally, the connections between observational data and implied causal connections between factors in statistical approaches involving regression in PER have been criticized. In particular, to select good models from several statistically equivalent ones (Lee and Hershberger, 1990), researchers should consider different aspects of the models when selecting the best model. For example, the potential instructional implications of each model should be considered, i.e., whether these instructional implications will have a positive influence on instructors and their pedagogical approaches. When a model is framed in this way, it can empower instructors so that they adopt effective practices and understand their role in recognizing and empowering students and affirming their work. Moreover, researchers should consider whether the instructionally beneficial models are also supported by additional evidence. This evidence may include but is not limited to researchers' own interview data or findings from prior studies. In other words, researchers should generate at least a few substantively meaningful different equivalent versions and deliberate based upon instructional implications and other evidence for why the proposed model is better than the others (Kline, 2015).

Here, we illustrate this point with quantitative data from a motivational survey administered to students in the second semester of the introductory physics course for bioscience majors. We use the physics identity framework (as explained in section IV) as an example to show how the proposed approach is used to select a good model from several statistically equivalent ones. To quantify the significance and relative strength of our framework links, we used structural equation modeling (SEM) (R Core Team, 2013). The models predicted students' physics identity through self-efficacy, interest, and perceived recognition. While each model is statistically equivalent, the instructional implications of each model are different. We initially tested gender moderation between different constructs using multi-group SEM (between male and female students) to investigate whether the relationship between the different motivational constructs was different across genders. There were no group differences at the level of weak and strong measurement invariance and the level of regression coefficients. Therefore, we proceeded to gender mediation analysis to understand how self-efficacy, interest, and perceived recognition mediate the effect of gender on physics identity at the end of the second introductory physics course for bioscience majors. There are 27 statistically equivalent models with different predictive relations between the three mediating constructs (self-efficacy, interest, and perceived recognition) (Lee and Hershberger, 1990). Here we discuss four of the models to show how our framework could guide the selection of a good model if there is additional evidence, e.g., from individual interviews to support the model. In all models, the model fit indices indicate a good fit to the data (MacCallum *et al.*, 1996). All path analysis results of the models are shown in Fig. 2.3.

First, we consider model 1 in which there is no predictive relationship between self-efficacy, interest, and perceived recognition. Instead, there are covariances between each motivational factor. Figure 2.3(a) shows the path analysis results of this SEM model. Next, we consider model 2 where self-efficacy predicts interest and perceived recognition, and interest predicts perceived recognition [see Fig. 2.3(b)]. In model 3, interest predicts self-efficacy and perceived recognition, and self-efficacy predicts perceived recognition [see Fig. 2.3(c)]. Finally, in model 4, perceived recognition predicts self-efficacy and interest, and self-efficacy predicts interest [see Fig. 2.3(d)].

Since all four models in Fig. 2.2 are statistically equivalent, according to the theoretical framework, we now must consider the instructional implications of each model. In model 1, self-efficacy and interest

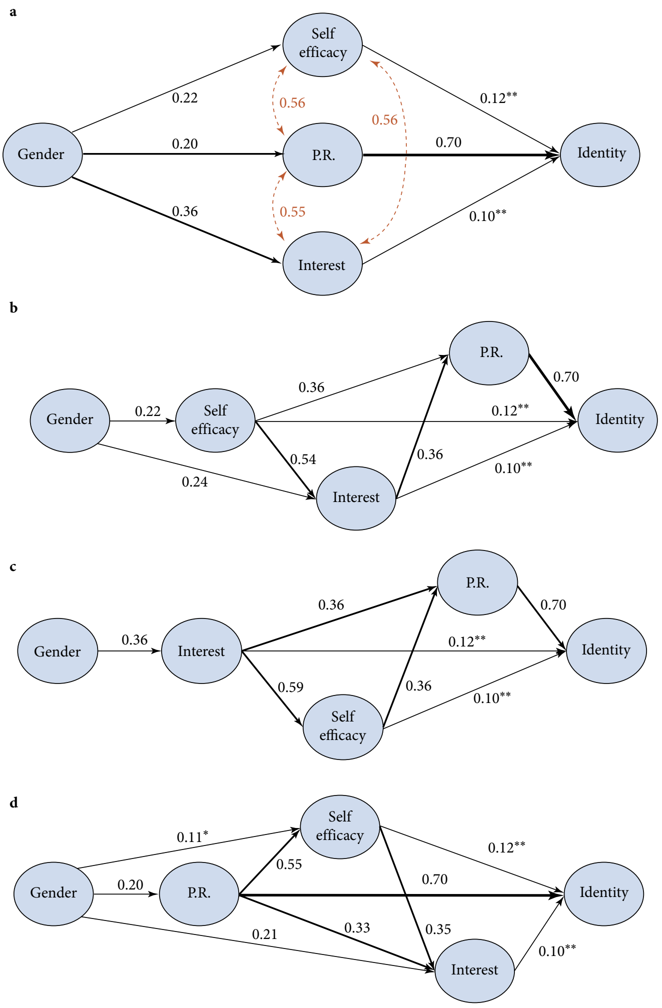


FIG. 2.3

Results of the path analysis part of SEM models 1–4 show how the relationship between gender and physics identity is mediated through self-efficacy, interest, and perceived recognition (P.R.). (a) In model 1, there are only covariances between each pair of constructs: self-efficacy, P.R., and interest. (b) In model 2, self-efficacy predicts interest and P.R., and interest predicts P.R. (c) In model 3, interest predicts self-efficacy and P.R., and self-efficacy predicts P.R. (d) In model 4, P.R. predicts interest and self-efficacy, and self-efficacy predicts interest. The dashed lines represent residual covariances between constructs. The solid lines represent regression paths, and the numbers on the lines are standardized regression coefficients (β values), which represent the strength of the regression relations. Each regression line thickness qualitatively corresponds to the magnitude of the β value. Regression coefficients with $p < 0.050$ are indicated by superscript ***, $p < 0.010$ are indicated by superscript **, and $p \leq 0.001$ are indicated by no superscript. For clarity, we have removed the statistically insignificant regression paths from gender to identity or gender to P.R. and self-efficacy in models 2 and 3.

are predicted only by gender. However, self-efficacy and interest may be considered fixed by physics educators and instructional policy makers. Model 1 does not provide suggestions that instructors can use to improve those motivational factors. Since perceived recognition, interest, and self-efficacy covary, it can be difficult for instructors to interpret how the motivational factors relate to one another. In addition, in models 2 and 3, gender only predicts self-efficacy or interest, respectively, which could be interpreted as a deficit model. In particular, these equivalent models can be interpreted to imply that women are not feeling positively recognized by their instructors and teaching assistants (TAs) as much as men because they have lower interest and self-efficacy than men. While statistically equivalent to models 1–3, model 4 with perceived recognition predicting self-efficacy and interest is more likely to give them the message that students' interest and self-efficacy in physics can be influenced by the recognition they receive from instructors and teaching assistants. Thus, model 4 is also more likely to inspire instructors to create a more inclusive and equitable learning environment in which all students, including those from the underrepresented groups, feel more positively recognized and affirmed.

Next, we must assess whether the instructionally beneficial models are also supported by additional evidence. Model 4 also supports and reflects findings from prior interviews. The interviews show that recognition by others, especially from instructors or TAs, is critical in shaping students' self-efficacy and interest (Doucette and Singh, 2020; Doucette *et al.*, 2020; Li *et al.*, 2020; and Santana and Singh, 2021). Thus, we argue that model 4 (in which physics self-efficacy, interest, and perceived recognition mediate the relation between gender and physics identity) is the best statistically equivalent model in models 2 and 4 are better than others based on our theoretical framework focusing on the model's instructional implications and supporting evidence from individual interviews with students.

In conclusion, to promote equity and inclusion, there is an urgent need to dismantle inequitable structures and create an equitable and inclusive learning environment. Our conception of equitable outcomes discussed earlier emphasizes that all demographic groups should have comparable outcomes. Taking inspiration from prior studies, an effective approach to creating an equitable and inclusive learning environment humanizes learning and takes advantage of student assets using culturally responsive pedagogy instead of using a deficit view of students (Kishimoto, 2018). In particular, institutions should recognize their responsibility with regard to taking action to create an equitable and inclusive learning environment and encourage the use of pedagogy in which all students have high motivational beliefs and can participate fully without the fear of being judged. It can also be beneficial for instructors to set an equity goal for their classes to explicitly track whether the demographic differences in outcomes in their courses are vanishing as they make them equitable and inclusive with a focus on student success. In addition, more research is needed at the intersection of various demographic factors (Cochran and Boveda, 2020; and Cochran *et al.*, 2020), and diverse selections of schools such as four-year institutions, community colleges, minority serving institutions, and all-women's colleges.

REFERENCES

Aguilar, L. *et al.*, Phys. Today **67**(5), 43–49 (2014).

AIP Statistics, see https://www.aip.org/statistics/data-graphics/percent-physics-bachelors-and-phds-earned-women-classes-1975-through-2016.

Archer, L. *et al.*, Am. Educ. Res. J. **54**(1), 88–126 (2017).

Bandura, A., Psychol. Rev. **84**(2), 191 (1977).

Bandura, A., *Encyclopedia of Psychology*, 2nd ed., edited by R. J. Corsini (Wiley, 1994), Vol. 3, pp. 368–369.

Barthelemy, R. S. and Knaub, A. V., Phys. Rev. Phys. Educ. Res. **16**(1), 010133 (2020).

Barthelemy, R. *et al.*, Phys. Rev. Phys. Educ. Res. **12**, 020119 (2016).

Binning, K. R. *et al.*, Psychol. Sci. **31**(9), 1059–1070 (2020).

Blickenstaff, J. C., Gender Educ. **17**(4), 369–386 (2005).

Blue, J. and Heller, P., AIP Conf. Proc. **720**, 45 (2004).

Blue, J. *et al.*, Phys. Today **71**(3), 40 (2018).

Boveda, M. and Weinberg, A. E., Phys. Teach. **58**, 480 (2020).

Brewe, E. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **6**(1), 010106 (2010).

Canning, E. A. *et al.*, Sci. Adv. **5**(2), eaau4734 (2019).

Carlone, H. B. and Johnson, A., J. Natl. Assoc. Res. Sci. Teach. **44**(8), 1187–1218 (2007).

Charleston, L. *et al.*, J. Progress. Policy Pract. **2**(3), 273–293 (2014).

Cho, S. *et al.*, Signs: J. Women Cult. Soc. **38**(4), 785–810 (2013).

Cochran, G. and Boveda, M., *PERC Proceedings* (PERC, 2020), pp. 9–15.

Cochran, G. *et al.*, *Handbook of Research in STEM Education*, edited by Routledge (2020), pp. 257–266.

Cochran, G. *et al.*, *Physics Education Research Conference*, Cincinnati, OH (2018), see https://scholar.google.com/citations?view_op=view_citation&hl=en&user=bOvomRsAAAAJ&citation_for_view=bOvomRsAAAAJ:LI9QrySNdTsC.

Coletta, V. P. *et al.*, AIP Conf. Proc. **1413**, 23 (2012).

Crenshaw, K., Stan. L. Rev. **43**, 1241 (1990).

Crowley, K. *et al.*, Psychol. Sci. **12**(3), 258–261 (2001).

Cwik, S. and Singh, C., Phys. Rev. Phys. Educ. Res. **17**(2), 020138 (2021a).

Cwik, S. and Singh, C., Phys. Rev. Phys. Educ. Res. **17**(2), 020143 (2021b).

Cwik, S. and Singh, C., Phys. Rev. Phys. Educ. Res. **18**(2), 020111 (2022a).

Cwik, S. and Singh, C., Phys. Rev. Phys. Educ. Res. **18**(1), 010138 (2022b).

Cwik, S. and Singh, C., Phys. Rev. Phys. Educ. Res. **18**(1), 010139 (2022c).

Cwik, S., "Investigating gender differences in students' motivational beliefs and inclusiveness of the learning environment in introductory physics courses for bioscience majors," Ph.D. Dissertation, University of Pittsburgh (unpublished, 2022d); available at http://d-scholarship.pitt.edu/43241/.

Cwik, S. *et al.*, *Physics Education Research Conference 2020* (PERC, 2020), see https://doi.org/10.1119/perc.2020.pr.Cwik.

Danielsson, A. T., Gender Educ. **24**(1), 25–39 (2012).

Dasgupta, N., Psychol. Inquiry **22**(4), 231–246 (2011).

Dew, M. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(1), 010106 (2021).

Doucette, D. *et al.*, Eur. J. Phys. **41**(3), 035702 (2020).

Doucette, D. and Singh, C., Phys. Teach. **58**(5), 297–300 (2020).

Doucette, D. and Singh, C., Phys. Teach. **60**(3), 166–168 (2022).

Dweck, C. S., *Mindset: The New Psychology of Success* (Random House Digital, Inc., 2008).

Espinosa, T. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(1), 010132 (2019).

Gonsalves, A. J. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020120 (2016).

Good, M. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020129 (2019).

Gutmann, B. and Stelzer, T., Phys. Rev. Phys. Educ. Res. **17**(2), 020121 (2021).

Hazari, Z. *et al.*, Phys. Teach. **55**(2), 96–99 (2017).

Hazari, Z. and Cass, C., Phys. Teach. **56**(7), 442–446 (2018).

Hazari, Z. *et al.*, J. Res. Sci. Teach. **57**(10), 1583–1607 (2020).

Hazari, Z. *et al.*, J. Res. Sci. Teach. **47**(8), 978–1003 (2010).

Hazari, Z. *et al.*, Sci. Educ. **91**(6), 847–876 (2007).

Henderson, R. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(2), 020103 (2018).

Henderson, R. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 020114 (2017).

Hyater-Adams, S. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(1), 010132 (2018).

Johnson, A., *Physics Education and Gender: Identity as an Analytic Lens for Research*, edited by A. J. Gonsalves and A. T. Danielsson (Springer International Publishing, 2020), pp. 53–80.

Johnson, A. *et al.*, Phys. Teach. **55**(6), 356–360 (2017).

Kalender, Z. Y. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020119 (2019a).

Kalender, Z. Y. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020148 (2019b).

Kalender, Z. Y. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010118 (2020).
Kalender, Z. Y. *et al.*, Phys. Rev. Phys. Educ. Res. **18**, 010116 (2022).
Kanim, S. and Cid, X. C., Phys. Rev. Phys. Educ. Res. **16**, 020106 (2020).
Karim, N. I. *et al.*, Eur. J. Phys. **39**(2), 025701 (2018).
Kelly, A. M., Phys. Rev. Phys. Educ. Res. **12**(2), 020116 (2016).
Kishimoto, K., Race Ethn. Educ. **21**(4), 540–554 (2018).
Kline, R. B., *Principles and Practice of Structural Equation Modeling* (Guilford Publications, 2015).
Kost, L. E. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **5**(1), 010101 (2009).
Kost-Smith, L. E. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **6**(2), 020112 (2010).
Ladewig, A. *et al.*, Front. Psychol. **11**, 548781 (2020).
Lee, S. and Hershberger, S., Multivariate Behav. Res. **25**(3), 313–334 (1990).
Lewis, K. L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020110 (2016).
Li, Y. and Singh, C., Phys. Rev. Phys. Educ. Res. **17**(1), 010143 (2021).
Li, Y. and Singh, C., Phys. Rev. Phys. Educ. Res. **18**(1), 010142 (2022).
Li, Y. *et al.*, Phys. Teach. **58**(7), 484–487 (2020).
Lindstrøm, C. and Sharma, M. D., Int. J. Innovation Sci. Math. Educ. **19**(2), 1–19 (2011).
Little, A. J. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(1), 010127 (2019).
Lorenzo, M. *et al.*, Am. J. Phys. **74**(2), 118–122 (2006).
MacCallum, R. C. *et al.*, Psychol. Methods **1**(2), 130 (1996).
Madsen, A. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **9**(2), 020121 (2013).
Malespina, A. *et al.*, Int. J. STEM Educ. **9**(1), 28 (2022).
Malespina, A. and Singh, C., Eur. J. Phys. **43**(3), 035701 (2022).
Marchand, G. C. and Taasoobshirazi, G., Int. J. Sci. Educ. **35**(18), 3050–3061 (2013).
Maries, A. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(2), 020119 (2018).
Maries, A. *et al.*, Phys. Teach. **58**, 430–433 (2020).
Maries, A. *et al.*, J. College Sci. Teach. **51**(3), 27–36 (2022).
Marshman, E. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(2), 020123 (2018a).
Marshman, E. *et al.*, Can. J. Phys. **96**(4), 391–405 (2018b).
McCullough, L., J. Int. Women's Stud. **5**(4), 20–30 (2004), see https://vc.bridgew.edu/jiws/vol5/iss4/2.
Mears, M., Phys. Rev. Phys. Educ. Res. **15**(2), 020135 (2019).
Mitchell, J. D. *et al.*, *Intersectionality & Higher Education* (Peter Lang, 2014).
Miyake, A. *et al.*, Science **330**(6008), 1234–1237 (2010).
Monsalve, C. *et al.*, *Proceedings of the Physics Education Research Conference* (PERC, Sacramento, CA, 2016).
Morton, T. R. and Parsons, E. C., Sci. Educ. **102**(6), 1363–1393 (2018).
Nissen, J. M. and Shemwell, J. T., Phys. Rev. Phys. Educ. Res. **12**(2), 020105 (2016).
Ong, M., Soc. Probl. **52**(4), 593–617 (2005).
Pollack, E., *The Only Woman in the Room: Why Science is Still a Boys' Club* (Beacon Press, 2015).
Pollock, S. J. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **3**(1), 010107 (2007).
Quichocho, X. R. *et al.*, *2019 PERC Proceedings* (PERC, 2019), pp. 24–25.
Quinn, K. N. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010129 (2020).
R Core Team (2013), see https://www.r-project.org/.
Rodriguez, I. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **8**(2), 020103 (2012).
Rodriguez, I. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020118 (2016).
Rodriguez, M. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 013101 (2022).
Rosa, K. and Mensah, F. M., Phys. Rev. Phys. Educ. Res. **12**(2), 020113 (2016).
Rosen, D. J. and Kelly, A. M., Phys. Rev. Phys. Educ. Res. **16**(2), 020116 (2020).
Salehi, S. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020114 (2019).
Santana, L. M. and Singh, C., *Physics Education Research Conference 2021* (PERC, 2021), see https://www.compadre.org/per/items/detail.cfm?ID=15784.
Sawtelle, V. *et al.*, J. Res. Sci. Teach. **49**(9), 1096–1121 (2012).
Steele, C. M. and Aronson, J., J. Personality Soc. Psychol. **69**(5), 797–811 (1995).
Stewart, J. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(1), 010107 (2021).
TEAM-UP Report, see https://www.aip.org/sites/default/files/aipcorp/files/teamup-full-report.pdf (2020).
Traxler, A. L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020114 (2016).
Traxler, A. L. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(1), 010103 (2018).
Van Dusen, B. and Nissen, J., J. Res. Sci. Teach. **57**(1), 33–57 (2020).
Wang, J. and Hazari, Z., Phys. Rev. Phys. Educ. Res. **14**(2), 020111 (2018).
Watkins, J. E., *Examining Issues of Underrepresented Minority Students in Introductory Physics* (Harvard University, Cambridge, MA, 2010).
Whitcomb, K. M. *et al.*, AERA Open **7**(1), 1–16 (2021).
Whitcomb, K. M. *et al.*, Int. J. Eng. Educ. **36**(6), 1996–2014 (2020), see https://www.ijee.ie/1atestissues/Vol36-6/24_ijee4004.pdf.
Whitcomb, K. M. and Singh, C., Eur. J. Phys. **41**(6), 065701 (2020).
Whitcomb, K. M. and Singh, C., Int. J. Sci. Educ. **43**(7), 1054–1089 (2021).

CHAPTER

3 RESEARCH ON GENDER, INTERSECTIONALITY, AND LGBTQ+ PERSONS IN PHYSICS EDUCATION RESEARCH

Ramón S. Barthelemy, Adrienne L. Traxler, Jennifer Blue, and Madison Swirtz

Barthelemy, R. S., Traxler, A. L., Blue, J., and Swirtz, M., "Research on gender, intersectionality, and LGBTQ+ persons in physics education research," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 3-1–3-16.

LGBTQ+: *An acronym that stands for Lesbian, Gay, Bisexual, Transgender, and Queer, which represents the diverse spectrum of people who identify their sexuality and gender differently than our societally expected identities based on culture and assignment at birth.*

3.1 INTRODUCTION

Physics Education Research (PER), in recent years, has begun to seriously consider the challenges of underrepresented and marginalized persons in physics (Barthelemy *et al.*, 2016; Rosa and Mensah, 2016; Traxler *et al.*, 2016; Hyater-Adams *et al.*, 2018; Hyater-Adams *et al.*, 2019; Barthelemy, 2020; Quichocho *et al.*, 2020; and Rodriguez *et al.*, 2022). Historically, the field began to address issues of gender, as defined by the binary, in the early 1990s, with discussions of race largely being taken up in the following decades. However, conversations at the intersections of these identities, or considering others such as being LGBTQ+ or a person with disabilities, have been limited (Barthelemy, 2020; James *et al.*, 2020; and Quichocho *et al.*, 2020). The studies that have been conducted have largely relied on deficit comparisons where a marginalized group is compared to a group in the majority (i.e., women compared to men, or People of Color compared to white people) (Scherr, 2016; and Traxler *et al.*, 2016). Recent suggestions have pointed to the limited capacities of such comparisons, and to the real world realities that representation and experiences of people from marginalized groups have remained relatively unchanged over the last few decades (Traxler *et al.*, 2016; and Rodriguez *et al.*, 2022).

https://doi.org/10.1063/9780735425514_003

The International Handbook of Physics Education Research: Special Topics, Mehmet Fatih Taşar and Paula R. L. Heron

This chapter seeks to understand, and gently critique, the work that has been done on women and LGBTQ+ persons in physics with attention paid to the intersections of race within these identity categories. Section 3.2 will explore gap-gazing, the phenomena of scholars focusing on the differences between marginalized and majority groups. Here, trends across research will be discussed with examples of conducting anti-gap work. Section 3.3 will zoom in on intersectionality and how work at the crossroads of race and gender is crucial for unpacking the lived experiences of Women of Color in physics and STEM. Section 3.4 will be centered on the experiences of LGBTQ+ persons in physics and STEM, an emerging area of research focus.

3.2 AVOIDING GAP-GAZING

What is gap-gazing, and why is it bad? For the purposes of this chapter, "gap-gazing" refers to research that focuses on quantitative, gender-based performance gaps to the exclusion of other questions and answers. Similar work has been done in the literature on other marginalized groups, such as People of Color. More attention was given to this intersection in Chap. 2. The last part of the above gap-gazing definition, the exclusion of other questions and answers, is the real problem. A great deal of important work has focused on performance gaps, but the space we can explore is limited if that is the only road that researchers follow. In this section, we will explore the different kinds of studies represented across gender gap work.

Quantitative work tends to be favored in PER because many researchers come from physics backgrounds where they were trained to frame investigations in quantitative terms. This is a powerful and productive way of understanding the world, but it is not equally suited to all research questions. Complicating matters, this quantitative bias does not just come from training but also from a need to be perceived as doing "real physics" in order to belong (and stay employed in) physics departments.[1] Thus, it stems at least partly from the social system of science and academia, rather than from what research questions are the most wide-reaching or most important to pursue. In brief, if it can be measured using a multiple-choice test, it is much more likely to be studied and published in PER than if it can only be studied through deep qualitative analysis. This is a serious limitation of the field.

Gender-based studies in PER have largely focused on contrasting test scores between men and women. These tests might be class exams or standardized instruments such as the Force Concept Inventory (Hestenes *et al.*, 1992) or the Colorado Learning Attitudes about Science Survey (Adams *et al.*, 2006). It is much harder to find studies on gender effects in class participation, social roles in groups, faculty responses to questions, or other areas where gender might matter in the learning environment (Danielsson, 2010). Additionally, most studies treat gender and sex as synonymous, reporting gender as a binary (Traxler and Blue, 2020) and often assigning it from school records (which may not match students' identities). The possibility of students with non-binary genders is not allowed in this research approach (Rasmussen *et al.*, 2019).

---

[1] This need can be understood as a kind of boundary work (Gieryn, 1983).

Performance gaps are important to identify because they often signal problems in the educational environment (Lubienski, 2008) or gaps in resource allocation. These problems are difficult to discuss or remedy without naming and describing them. However, the framing of these conversations matters. If all measurements are of individual traits, for example, the social dynamics of groups or instructors may go unstudied. If students are the only objects of scrutiny, sexist or racist structures in the learning environment may also go unrecorded. An example is work on stereotype threat by Claude Steele and others (Steele and Aronson, 1995; and Aronson, 2004), where significant differences in test performance by groups from marginalized backgrounds could be lessened or removed by manipulating the context of the test-taking. These effects were only discovered when the scope of the inquiry was widened from students' scores to the surrounding social and psychological conditions.

There are many possible routes to avoid gap-gazing. Below, we discuss two examples: explicitly anti-gap work using alternate frameworks and comparison between groups of women or otherwise focusing only on women.

3.2.1 Explicitly anti-gap work

Some researchers address the problems of gap-gazing by consciously using anti-gap theoretical frameworks. One example is a study by Danielsson (2012) that constructs discourse models from case studies on how physics students at a Swedish university perceive identity tracks in physics. The work aims to look at how women "do physics" and negotiate gender, without falling back on the common stereotypes of men and women in physics. (Examples of such stereotypes are an opposition of interested and engaged men vs under-confident women who need the physics to seem more relevant to everyday life before they engage.) The theoretical framework is deliberately anti-essentialist (meaning identities are not simple, essential, and static categories), focusing on nuance and multiple possibilities within gender rather than building large, unified categories of what men and women are "like." This connects to calls (Gutiérrez, 2008) for more work that studies excellence within groups rather than comparisons.

Another example is a study by Hughes (2001) that presents a critical discourse analysis of student interview pairs at a London city school and post-16 college. The interviews show nuances of multiple subject positions adopted by students as well as field differences between biology and physics. The author concludes by arguing that constructivist and society-relevant science teaching are important even when avoiding gender essentialism because they create a wider space for students to construct science identities. This last point highlights one of the framework shifts permitted by qualitative work. In quantitative studies, there is often explicitly or implicitly a single correct set of answers or attitudes that students "should" have, and they are measured for their degree of alignment with this goal. Qualitative studies are more likely to present a wider range of trajectories that students might take (Brickhouse, 2001), which is a useful theoretical lens for researchers who want to broaden their participation in physics.

3.2.2 Comparing among women

Another alternative to binary comparisons of men and women is to foreground women's experiences in the study design and analysis. One example is the work by Gonsalves (2014) that uses embedded case studies to examine women doctoral students in a Canadian physics department. A key theme discussed in this study is recognition as an aspect of physics identity, achievable through demonstrations of competence or through performance of stereotypical discourses of physicists. Gonsalves' work highlights that the self-image of physics as "gender neutral" ignores the negotiation of gendered subject positions that women in the field have to make.

Another example is a study by McCormick *et al.* (2014) of successful women students in an astronomy Ph.D. program in the United States. This study also drew on work by Whitten and collaborators (2003; 2004; and 2007) that looked for common "what works" elements in departments in the United States that graduated many women with physics degrees. Whitten and collaborators found, and McCormick *et al.*'s work bore out, that mentoring experiences, peer support, and collaborative learning were recurring elements in these success stories. With the exception of collaborative learning, these elements are rarely the focus of "gap" studies, so would be unlikely to be observed there. McCormick *et al.* also found two additional elements common to their participants: a passion for astronomy and good access to research and educational opportunities, often from a young age.

A third example is work by Dabney and Tai (2014) that looks for differences in background and motivational factors among women who pursued careers in physics or chemistry. Their work is quantitative and draws on survey data from Project Crossover, a mixed-methods study in the United States. They found that physics career choice for women was associated with the highest physics grades in secondary school and university, with less-than-highest chemistry grades, and with good and bad experiences respectively in university physics and chemistry. These differences had not been noted in previous studies on gender and STEM career choice, which typically focus on comparing men and women.

3.2.3 Advantages and limitations of gap-based studies: The need to include diverse voices

The examples above show some of the nuances that can be uncovered by departing from gap-focused studies; more on these themes can be found in Blue *et al.* (2019). Summaries and exemplars of PER's prevailing tradition of gender research have been given recently by McCullough (2016; and 2018) and Danielsson (2010). The anti-gap studies we have focused on in this section do not invalidate work on common gendered trends, e.g., that physics' reputation for not benefiting society may be especially damaging among women students. Instead, they enrich and complicate it with links to the diversity of lived experience. There is a tension inherent in trying to honor individual differences and trying to understand—and make policy around—large-scale patterns. This tension can be difficult to navigate but is unavoidable and ultimately can benefit PER if it keeps researchers honest about the limitations of our work.

To note some of those limitations, work on "gender in physics" has primarily meant work on *women* in physics. This comes from a tendency to treat men as the default, and gender as something that mostly women have. This framing positions women (rather than the field) as the thing that needs to change, and it does not acknowledge the existence of other genders or the spectrum of gender expression even within the binary categories of "men" and "women." In the last decade, gender identity has become much more prominent in media and cultural conversations in many countries. We hope that this nuance will propagate into research study design (Fernandez *et al.*, 2016; and Rasmussen *et al.*, 2019).

Finally, studies of gendered achievement gaps in physics usually do not consider race or ethnicity, and in many cases this means that research on "women in physics" is implicitly research on *white* women in physics. This framing neglects to consider intersections of race and gender, which are complicated and powerful. We now turn to this question in more detail.

3.3 INTERSECTIONALITY

Patricia Hill Collins introduced the idea of a "matrix of domination" in her book *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment* when it was first published in 1990 (Collins, 2000). The matrix of domination describes systems in which different identities like race, sex, and class interlock to give people in one interlocking group (for example, Black women) a different experience than those in another group (like white women).

Kimberle Crenshaw, a law professor, is credited with inventing the term intersectionality in a 1989 paper (Crenshaw, 1989). There she showed that using what she calls a single-axis framework, looking at just race or just at gender, marginalizes and erases the experiences of Black women. She uses several law cases as examples of times where looking at discrimination against women (mostly white women) or Black people (mostly men) missed the experiences of Black women and did not end up treating them fairly. Crenshaw argues the need for an intersectional approach. In other words, analyses of racism must include a look at sexism, and feminism must include an analysis of race.

Researchers in science education have been called upon to keep their work intersectional, recently by Geraldine Cochran, Mildred Boveda, and Chanda Prescod-Weinstein in the 2020 *Handbook of Research on STEM Education* (Cochran *et al.*, 2020). They use both Critical Race Theory (CRT) and intersectionality to ask researchers to use intersectional perspectives, theories, and frameworks, giving multiple examples. Examples of such work and productive ways to move forward are presented below.

3.3.1 Intersectional work

Louise Archer, Jennifer DeWitt, and Jonathan Osborne studied children in the United Kingdom in their paper "Is science for us? Black students' and parents' views of science and science careers" (2015). They surveyed and interviewed the same participants three times, in Year 6 (ages 10 and 11), Year 8 (ages 12 and 13), and Year 9 (ages 13 and 14). In all three years, the survey results showed that Asian students had the most desire to study and work in science, followed by Black students and then white students. These differences were significant, though they had small effect sizes. Further statistical modeling showed that other factors were likely more important than ethnicity. These factors include parental attitudes toward science, the students' own attitudes toward science, and the students' gender.

Interviews were conducted in all three years with all the Black children in the sample: 10 children (four male and six female) and six of their parents. Of course, the authors acknowledge that there might be some bias here, as the families of children interested in science were probably more likely to volunteer for interviews.

Why did the majority of Black students not want STEM careers? They were good at science, interested in science, and wanted to do well in school. Although they agreed that math and English would be useful for most jobs, they could only see school science as useful for careers as a science teacher, scientist, or doctor. They could not imagine themselves in those careers, which they dismissed as boring. Furthermore, they thought people had to be exceptionally smart to be a scientist, plus geeky, and they did not see themselves there. The authors suggested that performing Black masculinity or femininity did not intersect with these traits. The Black parents also had fairly traditional views of gender and did not see science (or brilliance) as important for their daughters. Black parents in the study also did not think of themselves as knowing much science; it did not seem attainable to them or, in most cases, to their children.

There were two participants in the study who were low-income Black girls, both children of Nigerian immigrants, who wanted to be (forensic) scientists. They were inspired by the television programs *CSI* (*Crime Scene Investigations*), set in the United States, and *Death in Paradise*, set in the Caribbean. In their families, the performance of Black femininity was consistent with being quietly excellent students, as they were. Although the father of one of the girls worked as a scientist, their mothers still both nudged them toward jobs like doctor and pharmacist, perhaps worried that their daughters would be disappointed.

We can look at what happens to Women of Color in college in one of the earliest and most-cited articles on the intersection of race and gender in physics, Maria Ong's "Body projects of young Women of Color in physics: Intersections of gender, race, and science" (Ong, 2005). She starts with introductions on what it means to be "ordinary," or fit in, in the three realms of science, gender, and race/ethnicity, and finishes her literature review by addressing fragmentation, multiplicity, and body theory. Negotiating multiple identities sometimes leads to fragmentation or "passing" as people try to minimize identities that make them seem different from others. More rarely, this negotiation leads to multiplicity, in which people boldly embody all of their identities at the same time. Body theory emphasizes the human body as the vehicle through which this fragmentation or multiplicity is displayed; people might accentuate or de-emphasize their curves, for example.

Ong's study involved ten Women of Color, whom she interviewed every year, starting when they were in university and continuing over an eight-year period. Although all students earned an undergraduate degree in physics (or a closely related field like engineering physics), several notable themes emerged from their interviews. Their peers often assumed that they were not smart enough to do physics, at least until they joined research labs, got high grades, or were awarded fellowships, thus being publicly recognized for their work. She also noted that some students used fragmentation to their benefit; they passed as white and/or acted fairly masculine, while others used multiplicity to their benefit; they made themselves stand out by "performing" their gender and ethnicity.

Ong concludes by noting that the very intersectionality of being a Woman of Color put these students in a more precarious place than Men of Color or white women. Further, she notes that the students who fragmented themselves tended to get along better in their departments than students who used multiplicity, but this happened at a personal cost.

The obstacles facing People of Color continue through graduate school, as Brian A. Burt, Alexander Knight, and Justin Robeson wrote in "Racializing experiences of foreign-born and ethnically diverse Black male engineering graduate students: Implications for student affairs, policy, and research" (2017). The study is grounded in Critical Race Theory (CRT), accepting that race is a social construct and using the CRT tenet of anti-essentialism, asserting that there is intersectionality even within a racial group. Not all Black people, or even all Black men, are the same. The authors interviewed nine Black men attending graduate school in majority white engineering programs in the United States. All the men in the study held ethnic identities other than African American, and six of them were born outside the U.S.

Three major themes emerged from the interview analysis. One is racialization, as the Black men learn about American racism and how it will affect them. The other is cultural tensions, as the men note differences between themselves and people with different cultural and national origins. Some of them did not feel as though they belonged in the U.S. culture, including the Black U.S. culture, even if adopting that culture would increase their success in this country. The third theme is a particular kind of impostor syndrome. After already noting that they might not belong in U.S. culture, the subjects also wondered if they belonged in engineering. They were usually the only Black person in a room of engineering students (or engineers), and one noted that, "*When I meet someone, the chance of them—you know guessing that I am a Ph.D. student in engineering is zero*" (p. 936).

The authors conclude their paper with recommendations that university faculty and staff, particularly those involved in orientation and working with international students, expand their programming. Students need support in navigating norms, transitions to both the academic and cultural aspects of graduate school, and feelings of impostor syndrome.

Katemari Rosa and Felicia Moore Mensah grounded their paper, "Educational pathways of Black women physicists: Stories of experiencing and overcoming obstacles in life" (2016), in Critical Race Theory. They cite important work in CRT, highlighting three important aspects: (1) racism is permanent and seems natural in the culture of the United States; (2) it is necessary to engage in counter storytelling, highlighting the stories and lived experience of Black people; and (3) societal changes that benefit Black people will only happen when those changes also benefit white people.

Rosa and Mensah conducted long, semi-structured interviews of six Black women with Ph.D.s in physics, astronomy, or related fields (at a time when there were fewer than 100 people meeting that description in the United States). They chose participants who had most of their education in the United States; all identified as women, and they ranged in age from their twenties to their fifties. They analyzed the interview data with *in vivo* coding, allowing themes to emerge from their participants' words.

Some of the themes that emerged from the interviews that Rosa and Mensah conducted were negative. The participants had all experienced racism: teachers who dismissed their interest in science, countless instances of racial microaggressions, and exclusion from the study groups. The participants also felt isolated, both as women and as Black people.

Rosa and Mensah did also uncover some more positive themes. Many of the participants had early invitations to participate in science, several doing research as early as high school. They cited the feeling of being in a scientific community as well as the information they absorbed about being a working scientist. Every one of the participants also attended summer research programs while in college—and all of them got paid for that work. Several of the students also chose their majors, at least in part because there were scholarships available; four of the six participants mentioned scholarship packages in their interviews. It is also notable that four of the participants attended historically Black colleges and universities, whether for undergraduate or graduate school.

Studies whose primary focus is not intersectionality can also support this research value by noting the diversity of identities in their participants (Gandhi *et al.*, 2016). This diversity often also includes gender and sexual identity, the focus of Sec. 3.4.

3.4 LGBTQ+ RESEARCH IN STEM EDUCATION

LGBTQ+ identity is an even more recent addition to discussions of identity in physics than gender or race and ethnicity. Because of its newness, this area of the literature is still very underdeveloped globally, but there are already themes emerging that echo Secs. 3.2 and 3.3. It is useful to consider the LGBTQ+ rights trajectory in the United States as an example of growth in recent years. The rights of lesbian, gay, bisexual, transgender, and queer (LGBTQ+) people in the U.S. have been changing significantly over the last two decades. This is juxtaposed to a long history of discrimination, legalized persecution, the survival of a pandemic (HIV/AIDS) at first ignored by national leaders, and few legal rights or protections (Sullivan, 2003). Most recently, a combination of public sentiment, supreme court rulings, and dueling executive orders have reified LGBTQ+ rights in the U.S. In 2015, a landmark supreme court ruling, Obergefell v. Hodges, legalized same-sex marriage (Doherty *et al.*, 2019). This was followed by a 2020 supreme court ruling, Bostock v. Clayton County, which finally granted LGBTQ+ workers federal protections against discrimination (Valenti, 2020). However, these large accomplishments have not protected LGBTQ+ persons against discrimination in U.S. culture at large and discriminatory policies by sitting politicians (Johnston and Meyer, 2017; Goodwin and Chemerinsky, 2019; and Kaufman and Compton, 2020).

An example of continued discrimination can be seen in the 2017 banning of transgender troops serving in the U.S. military by a sitting president (Goodwin and Chemerinsky, 2019). Even with same-sex marriage recently in the rear-view mirror, and federal work protections on the horizon, the then leader of the USA openly and purposefully discriminated against trans persons with a single signature (executive order). Fortunately, this ruling was overturned in 2021 with another signature (Chang, 2021). Discrimination at the discretion of the president is problematic because rights can change suddenly and violently. The largest gained rights, too, were achieved through Supreme Court findings when the conservative majority was 5-4, compared to the current 6-3 conservative supermajority, which puts liberal causes such as LGBTQ+ rights and women's access to healthcare in danger (Benen, 2020). Furthermore, acceptance of the various subgroups of LGBTQ+ persons can be uneven (Lewis *et al.*, 2017). Challenges remain around the world for LGBTQ+ persons, from social shunning of individual identities to death sentences for queer people. Many of these laws also stem from the British colonial era (Wong, 2021).

The sciences, and physics specifically, have not been immune to this unevenness of support, and evidence of discrimination. Currently, about 15% of early career physicists identify as LGBTQ+ (The Greater US L.L.C and American Physical Society, 2018). In physics, overall LGBTQ+ representation is about 4.5% (Cech and Waidzunas, 2018). A growing body of literature has begun to emerge focusing on the lives and experiences of LGBTQ+ scientists (Bilimoria and Stewart, 2009; Cech and Waidzunas, 2011, 2018; Patridge *et al.*, 2014; Yoder and Mattheis, 2016; Barres *et al.*, 2017; Cech and Pham, 2017; and Mattheis *et al.*, 2019), with a smaller subset specifically focusing on physics (Atherton *et al.*, 2013, 2016; Barthelemy, 2020; and Barthelemy *et al.* 2022a, 2022b). This review will survey the broad findings from this body of work with specific examples from the literature.

3.4.1 LGBTQ+ persons in STEM

3.4.1.1 Students

The literature on LGBTQ+ STEM students is confined to a handful of articles as of January 2020 (Cech and Waidzunas, 2011; Cooper and Brownell, 2016; Stout and Wright, 2016; Hughes, 2018; Linley *et al.*, 2018; and Freeman, 2020). The first such article came from Cech and Waidzunas (2011) and focused on the issues faced by lesbian, gay, and bisexual engineering students. Their study employed interviews and focus groups with 17 engineering students and sought to understand the climate they experienced and their methods of coping. Their findings indicated the necessity of students to hide their sexual orientation to a heteronormative climate (one that assumes heterosexuality). Their necessity to hide their identities was suggested to come with high emotional labor and anxiety about future job prospects and potential loneliness.

Another article on LGBTQ+ students came out seven years later, wherein a sample of over 4000 STEM-aspiring students it was demonstrated that LGBTQ+ students were less likely to be retained in their majors (Hughes, 2018). Further work from 2016 to 2020 pointed to the desires of students to have more out faculty and concerns of coming out themselves (Cooper and Brownell, 2016), perceived discrimination in larger STEM community and acceptance in peer groups (Linley *et al.*, 2018), low sense of belonging (Stout and Wright, 2016), and the need for more census data on LGBTQ+ persons (Freeman, 2020). Freeman (2020) specifically pointed out that LGBTQ+ persons are underrepresented in STEM but lack the resources given to other underserved groups. Combined, these articles describe a STEM environment for LGBTQ+ students, which is dismissive of their identities and impacts their ability to be fully a part of their STEM communities. Research on STEM faculty, students and the workforce also points to such trends (e.g., Barthelemy *et al.*, 2016). For example, in Barthelemy *et al.* (2016), the authors found that women in graduate physics and astronomy face a myriad of barriers, which unfortunately included discriminatory treatment by their peers based on their identity as women.

3.4.1.2 Workforce

The first qualitative study on STEM faculty appeared in 2009, followed by a quantitative study in 2014 (Bilimoria and Stewart, 2009; and Patridge *et al.*, 2014). Bilimoria and Stewart (2009) found in their study of 14 LGBTQ+ STEM faculty that they often felt invisible in their overwhelmingly heterosexual STEM communities. In addition to this invisibility, they experienced direct and indirect hostility toward them for being LGBTQ+, which resulted in them being excluded from certain professional networks, which ultimately impacted their careers.

Patridge *et al.* (2014) conducted a quantitative study building on previous 2010 work by analyzing survey responses from 279 LGBQ (queer) STEM faculty in the USA. Their study looked at the factors influencing their participants' climate experiences in STEM. The authors found a correlation between outness and being uncomfortable as well as the increased likelihood of STEM faculty leaving their jobs if they experienced or observed exclusionary behavior. Only about 10% of this sample reported being out about their LGBQ identity in their workplace. Work by Yoder and Mattheis (2016) added more nuance to this conversation by showing that STEM faculty were less likely to be out in professional settings as compared to personal settings in a survey of over 1400 participants. Qualitative work has continued to demonstrate the issues STEM faculty face in coming out (Barres *et al.*, 2017; and Cooper *et al.*, 2019).

Larger studies on LGBTQ+ STEM professionals have been conducted in recent years. A survey of over 30 000 U.S. federal STEM professionals found negative workplace climates of LGBTQ+ persons, which were not mitigated by employment status (Cech and Pham, 2017). Another study by this same team analyzed survey results of over 25 000 STEM professionals and found similarly negative results (Cech and Waidzunas, 2021). Their findings indicated that LGBTQ+ STEM professionals were more likely to experience career limitations, harassment, and professional devaluation than their non-LGBTQ+ peers.

These participants also reported higher levels of considering leaving their jobs. All of these findings stood even when controlling for participant demographics (Cech and Waidzunas, 2021). Lastly, more theoretical work has suggested a model for queer identity in the workplace (Yoder and Mattheis, 2016). Mattheis *et al.* (2019) argued for three components of queer (or LGBTQ+) identity. These were defining one's queer identity, forming a professional identity in STEM and navigating these identities in the workplace (Yoder and Mattheis, 2016). Together, the view of LGBTQ+ faculty has similar constraints and negative outcomes as those faced by students. The work of Yoder *et al.* (2016) is not the only theoretical piece on LGBTQ+ issues in STEM. Two other papers have also considered this topic (Gunckel, 2009; and Heimlich, 2019).

3.4.2 LGBTQ+ persons in physics

The work on LGBTQ+ physicists is even more limited than STEM at large. In total, four pieces could be found in 2020 (Atherton *et al.*, 2016; Dyer *et al.*, 2019; Barthelemy, 2020; and Quichocho and Close, 2020) in addition to an online best practices guide for LGBTQ+ inclusion physics and astronomy (Ackerman *et al.*, 2018). Of the four pieces, two are member society surveys, the first of which was released by the American Physical Society in 2016 (Atherton *et al.*, 2016).

The LGBT Climate in Physics Survey (Atherton *et al.*, 2016) was the culmination of grassroots LGBTQ+ physicists working in collaboration with the American Physical Society to investigate LGBTQ+ representation in physics and recommend specific changes in policy and practices to create a more inclusive environment for LGBTQ+ physicists. Through online snowball sampling, they surveyed 324 physicists on their experiences. They found that the experiences of LGBTQ+ physicists varied greatly between identities; transgender and gender non-conforming participants were more than twice as likely to report experiencing exclusionary behavior than cisgender participants, with many pointing toward a lack of institutional support through healthcare, name changes, and the use of correct pronouns. LGBTQ+ women were also more likely to experience exclusionary behavior than LGBTQ+ men and some qualitatively discussed fetishization and stereotyping of their identities to be multiplicative such that they could not distinguish what part of the exclusion was based in misogyny and which was based in homophobia. LGBTQ+ People of Color also found that they experienced intersectional discrimination, although this qualitatively manifested itself as participants feeling that it was more difficult to come out or express LGBTQ+ identity because they were already experiencing discrimination on the basis of race and did not want to exacerbate that. Further peer reviewed publications on this survey have now been published as well (Barthelemy *et al.* 2022a, 2022b). Importantly, Barthelemy *et al.* (2022a) conducted more static analyses, which showed a lower likelihood of BIPOC LGBTQ+ to be out, which was supported by their qualitative responses explaining barriers of racism. In Barthelemy *et al.* (2022b), the authors reported on statistical models that showed a positive workplace climate was more predictive of retention than a negative one was of consideration to leave.

Barthelemy (2020) analyzed the qualitative responses to an open-ended question about exclusionary experiences from the APS report. While this was not intended to give a general overview of all issues

LGBTQ+ physicists face, it elaborates on the ways in which more common types of exclusionary behaviors manifest themselves. This study found that transgender people experienced unintentional and targeted harassment on two main fronts: gender expression and misgendering. Discrimination based on gender expression was also reported by cisgender people who presented in non-normative ways, with one participant reporting snide remarks from colleagues based on his choice to wear nail polish. Women reported on the frequency of sexist comments, with one participant saying she was only out to those she trusts because she already experiences a large amount of gendered microaggressions. LGB respondents also reported being asked inappropriate or overly personal questions about their sexuality, then having those colleagues designate those issues as unimportant or not real. Some respondents also reported sexual assault and stalking behavior from colleagues.

The Institute of Physics, Royal Astronomical Society, and Royal Society of Chemistry also created a report inspired by the LGBT Climate in Physics Report (Dyer *et al.*, 2019). Their survey consisted of about 1000 participants across the physical sciences (mainly physics, astronomy, and chemistry) in the United Kingdom and Ireland. This report also found uneven comfort and institutional support based on gender identity; transgender participants in this study were also twice as likely to experience exclusionary and harassing behavior compared with cisgender participants. Additionally, 44% of gay and 38% of lesbian participants were out to everyone at work, while only 21% of queer and questioning and 14% of bisexual participants were out at work. They also noted that "outness" at work was very context driven, as the nature of international travel and collaboration meant traveling to countries with varying levels of LGBTQ+ support and criminalization. Between job types, teachers and scientists working outside a university setting also had fewer opportunities for LGBTQ+ networking and connection, formally and informally, than their peers working and studying at universities.

Quichoho *et al.* (2020) aimed to understand physics identity development through the lens of Black, Indigenous, and Women of Color (BIWOC) in addition to LGBTQ+ women (who may also identify as BIWOC in this study). Eight participants were interviewed about their personal narratives and intersecting identities in and out of physics spaces. There were three main categories of responses: the cost of identity fragmentation, integrated performance of being a physicist, and the manipulation of that performance. Respondents talked about the "circus-freak" effect of behavior where coming out and existing with multiple minoritized identities in physics led to peers distancing themselves and acting generally uncomfortable in their presence. Participants also noted integrating their physics identity: since they did not fit the stereotypical mold of being a physicist, their physics identities were intentionally constructed rather than assumed. By doing this integration, participants were also better able to recognize the strengths and physics identities of their peers. Some participants also noted intentionally manipulating their identity presentation to align more with the expectations of physicists. One participant stopped wearing makeup and put-together outfits so her peers would take her more seriously, and then went back to wear makeup once she felt her peers knew she was smart. These strategies echo those discussed earlier for women doctoral students (Gonsalves, 2014) and Women of Color in physics (Ong, 2005). The paper also ends with a call to action to educate ourselves about

Black issues and support Black students and colleagues in physics by developing anti-racist, anti-misogynistic, and anti-homophobic policies in our own institutions.

Overall, these papers point toward an environment in physics which encourages the suppression and obfuscation of identities in order to be considered a "real" scientist. On top of the already challenging experience of STEM undergraduate, graduate, and professional environments, LGBTQ+ scientists must constantly evaluate the costs and benefits of being out, expressing themselves fully, and participating in advocacy work without endangering their careers or physical safety. This research also points toward the importance of consistently using intersectional frameworks to avoid unintentionally flattening LGBTQ+ identity and normalizing the experiences of white gay and lesbian scientists. This was particularly important for LGBTQ+ physicists who were also women and/or BIPOC.

3.5 CONCLUSION

In the last decade or two, the concept of gender in physics education research has greatly expanded. There are now both qualitative and quantitative studies from a range of perspectives, including intersectional work with race, ethnicity, and LGBTQ+ identity. These latter kinds of studies are still quite rare but vitally important. Throughout this chapter, the value of intersectional anti-gap research on issues of gender, race and LGBTQ+ persons in physics has emerged. From this perspective, the importance of investigating the experiences of people from marginalized backgrounds without comparisons to their majority peers is emphasized. This kind of research is crucial in uncovering the lived experiences of students, staff and faculty. Through these kinds of investigations, scholars can build new knowledge that can be used in producing best practices for inclusion. In anti-gap studies, it is important to understand intersecting identities of race, gender, being LGBTQ+ and more. Physicists, like all people, do not live one-dimensional lives; their lives are complicated and multidimensional (Ferguson, 2018) and should be factored into scholarship. Furthermore, the reader should take to heart the importance of finding underserved groups and conducting scholarship and advocacy to support them. As physics classrooms and communities continue to diversify, it is critical that PER follows suit.

REFERENCES

Ackerman, N. *et al.*, arXiv:1804.08406 (2018).
Adams, W. K. *et al.*, Phys. Rev. Spec. Top. – Phys. Educ. Res. **2**(1), 010101 (2006).
Archer, L. *et al.*, Sci. Educ. **99**(2), 199–237 (2015).
Aronson, J., Educ. Leadership **62**(3), 14–19 (2004).
Atherton, T. J. *et al.*, *LGBT Climate in Physics: Building an Inclusive Community* (American Physical Society, 2016).
Atherton, T. *et al.*, see http://lgbtphysicists.org/files/BestPracticesGuide.pdf for 2013.
Barres, B. *et al.*, Genome Biol. **18**, 62 (2017).
Barthelemy, R., Eur. J. Phys. **41**(6), 065703 (2020).
Barthelemy, R. S. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 010147 (2022b).
Barthelemy, R. S. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020119 (2016).

Barthelemy, R. S. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 010124 (2022a).
Benen, S., MSNBC (2020). See https://www.msnbc.com/rachel-maddow-show/reckoning-6-3-conservative-majority-supreme-court-n1244947
Bilimoria, D. and Stewart, A. J., Natl. Women's Stud. J. **21**(2), 18 (2009).
Blue, J. *et al.*, Am. J. Phys. **87**(8), 616–626 (2019).
Brickhouse, N. W., J. Res. Sci. Teach. **38**(3), 282–295 (2001).
Burt, B. A. *et al.*, J. Int. Stud. **7**(4), 925–943 (2017).
Cech, E. and Pham, M., Soc. Sci. **6**(12) (2017).
Cech, E. A. and Waidzunas, T. J., Eng. Stud. **3**(1), 1–24 (2011).
Cech, E. and Waidzunas, T. J., *STEM Inclusion Study Organization Report: APS* (University of Michigan, Ann Arbor, MI, 2018).
Cech, E. and Waidzunas, T. J., Sci. Adv. **7**(3), 3 (2021).
Chang, A., NPR (2021); see https://www.npr.org/2021/01/25/960465926/biden-ends-ban-on-trans-people-serving-openly-in-the-military.
Cochran, G. L. *et al.*, *Handbook of Research on STEM Education*, edited by C. C. Johnson *et al.* (Routledge, 2020).
Collins, P. H., *Black Feminist Thought Knowledge, Consciousness, and the Politics of Empowerment*, 2nd ed. (Routledge, 2000).
Cooper, K. M. and Brownell, S. E., CBE Life Sci. Educ. **15**(3), 1–19 (2016).
Cooper, K. M. *et al.*, J. Women Minor. Sci. Eng. **28**, 25 (2019).
Crenshaw, K., *A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics* (The University of Chicago Legal Forum, 1989), pp. 139–167.
Dabney, K. P. and Tai, R. H., Phys. Rev. Spec. Top. – Phys. Educ. Res. **10**(1), 010104 (2014).
Danielsson, A. T., *Never Mind the Gap! Gendering Science in Transgressive Encounters*, edited by M. Blomqvist and E. Ehnsmyr (University Printers, Uppsala, 2010), pp. 65–83.
Danielsson, A. T., Gender Educ. **24**(1), 25–39 (2012).
Doherty, C. *et al.*, see https://www.pewresearch.org/politics/2019/05/14/majority-of-public-favors-same-sex-marriage-but-divisions-persist/ for 2019.
Dyer, J. *et al.*, IOP (2019); see https://www.iop.org/about/publications/exploring-the-workplace-for-lgbtplus-physical-scientists.
Ferguson, R. A., *One-Dimensional Queer* (John Wiley & Sons, 2018).
Fernandez, T. *et al.*, 2016 ASEE Annual Conference and Exposition (2016), p. 14546.
Freeman, J. B., Policy Insights Behav. Brain Sci. **7**(2), 141–148 (2020).
Gandhi, P. R. *et al.*, Am. J. Phys. **84**(9), 696–703 (2016).
Gieryn, T. F., Am. Sociol. Rev. **48**(6), 781–795 (1983).
Gonsalves, A. J., Cult. Stud. Sci. Educ. **9**, 503–521 (2014).
Goodwin, M. and Chemerinsky, E., Northwestern Univ. Law Rev. **114**, 751 (2019).
Gunckel, K. L., J. Curric. Theor. **25**(2), 13 (2009).
Gutiérrez, R., J. Res. Math. Educ. **39**(4), 357–364 (2008).
Heimlich, J. E., *STEM of Desire* (Brill Sense, 2019), pp. 161–176.
Hestenes, D. *et al.*, Phys. Teacher **30**(3), 141–158 (1992).
Hughes, G., Gender Educ. **13**(3), 275–290 (2001).
Hughes, B. E., Sci. Adv. **4**(3), eaao6373 (2018).
Hyater-Adams, S. *et al.*, Phys. Rev. Phys. Educ. Res. **14**, 010132 (2018).
Hyater-Adams, S. *et al.*, Phys. Rev. Phys. Educ. Res. **15**, 020115 (2019).
James, W. *et al.*, Phys. Rev. Phys. Educ. Res. **16** (2020).
Johnston, T. R. and Meyer, H., Housing Care Support 20(3), 121–127 (2017).
Kaufman, G. and Compton, D. L., Sexual. Res. Soc. Policy 18, 321–330 (2020).
Lewis, D. C. *et al.*, Polit. Res. Quarterly **70**(4), 861–875 (2017).
Linley, J. L. *et al.*, J. Women Minor. Sci. Eng. **24**(1), 1–16 (2018).
Lubienski, S. T., J. Res. Math. Educ. **39**(4), 350–356 (2008).
Mattheis, A. *et al.*, J. Homosexuality **67**(13), 1839–1863 (2019).
McCormick, M. *et al.*, J. Women Minor. Sci. Eng. **20**(4), 317–340 (2014).
McCullough, L., *Women and Physics* (Morgan and Claypool Publishers, 2016).
McCullough, L., *Getting Started in Physics Education Research*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, 2018), Vol. 2.
Ong, M., Soc. Probl. **52**(4), 593–617 (2005).
Patridge, E. *et al.*, J. Women Minor. Sci. Eng. **20**(1), 75–98 (2014).
Quichocho, X. R. *et al.*, paper presented at the Physics Education Research Conference Proceedings, 2020.
Rasmussen, K. C. *et al.*, arxiv:1907.04893 (2019).
Rodriguez, Z. *et al.*, Phys. Rev. Phys. Educ. Res. **18** (2022).
Rosa, K. *et al.*, Am. J. Phys. 89, 751 (2021).
Rosa, K. and Mensah, F. M., Phys. Rev. Phys. Educ. Res. **12**(2), 020113 (2016).
Scherr, R., Phys. Rev. Spec. Top. – Phys. Educ. Res. **12**(2), 020003 (2016).
Steele, C. M. and Aronson, J., J. Personality Soc. Psychol. **69**(5), 797–811 (1995).
Stout, J. G. and Wright, H. M., Comput. Sci. Eng. **18**(3), 24–30 (2016).

Sullivan, N., *A Critical Introduction to Queer Theory* (NYU Press, New York, 2003).
The Greater US L.L.C, & American Physical Society, Early Career Physicists and APS (2018).
Traxler, A. and Blue, J., *Physics Education and Gender: Identity as an Analytic Lens for Research*, edited by A. J. Gonsalves and A. T. Danielsson (Springer International Publishing, 2020), pp. 129–152.
Traxler, A. L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2016).
Valenti, A., Employee Respons. Rights J **33**, 3–23 (2020).
Whitten, B. L. *et al.*, J. Women Minor. Sci. Eng. **13**(1), 37–76 (2007).
Whitten, B. L. *et al.*, J. Women Minor. Sci. Eng. **9**(3–4), 20 (2003).
Whitten, B. L. *et al.*, J. Women Minor. Sci. Eng. **10**(3), 229–242 (2004).
Wong, T., BBC News, 29 June 2021. See https://www.bbc.com/news/world-asia-57606847
Yoder, J. B. and Mattheis, A., J. Homosexuality **63**, 1–27 (2016).

CHAPTER 4

RESEARCH ON EQUITY IN PHYSICS GRADUATE EDUCATION

Diana Sachmpazidi

Sachmpazidi, D., “Research on equity in physics graduate education,” in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 4-1–4-10.

4.1 INTRODUCTION

American women, African Americans, Latinx Americans, and Native Americans have historically been denied access to education as well as political and economic participation in the United States (U.S.) (Oakes, 1990). The results of this historical exclusion are, in part, evident in the underrepresentation of these groups in STEM graduate education. In physics, in particular, women students continue to be significantly underrepresented in doctoral programs compared with men students. Although half of the U.S. population are women (United States Census Bureau, 2019), only 19% of all physics students are women (Mulvey *et al.*, 2019). Individuals from marginalized racial and ethnic communities in the USA (African American, Hispanic American, and Native American) comprise about 33% of the country’s population (United States Census Bureau, 2019). However, students from marginalized backgrounds earn less than 3% of the awarded doctoral degrees in physics. Participation in physics graduate education by underrepresented U.S. students has slightly increased for Hispanic/Latinx students, while Black/African American students remain significantly underrepresented (Mulvey *et al.*, 2019). The representation of U.S. students from historically excluded racial and ethnic communities continues to be low in terms of doctoral programs in physics and in the doctoral award recipients (Mulvey *et al.*, 2019).

Students from groups traditionally marginalized in STEM disciplines (Black, Latinx, and Indigenous students of all genders) experience the impacts of structural and institutional racism and sexism in U.S. universities. For example, marginalized students are often portrayed as “unfit” in STEM disciplines as intellectually inferior to white and Asian American men (McGee *et al.*, 2022). Such structural and cultural barriers are reflected, among others, in institutional actors’ deficit views around the science competency of marginalized students (Bensimon, 2005; and Ghee *et al.*, 2016) and result in continuing

https://doi.org/10.1063/9780735425514_004

to nurture the dominant stereotype of white maleness in science (Smith *et al.*, 2019). For example, while socialization in the graduate program through peer interactions is one critical factor associated with higher retention in physics (Sachmpazidi, 2021), it is found that Black women face exclusion from participation in peer study groups—a crucial academic support (Rosa and Mensah, 2013). Furthermore, the resources and capital that these students bring to their classes and the success they experience often go ignored and/or undervalued (Yosso, 2016; and Sabella *et al.*, 2017).

While research on the issue of inclusion and diversity has recently expanded in physics education research (PER), mainly focusing on gender issues and race, little work has been done to understand the experiences of other marginalized groups (people with disabilities, gender and sexual minorities) (Barthelemy *et al.*, 2022). Moreover, even less work has been done focusing on the experiences of these marginalized groups in physics graduate education. There is well-documented evidence that LGBTQ+ (Lesbian, Gay, Bisexual, and Transgender people; the + is used to represent the many other gender and sexual minority identities part of the LGBT+ identity who are not L, G, B, or T) physicists encounter exclusionary behavior (sexual and verbal harassment, homophobic comments, purposeful misidentification of gender, stereotyping, and expectations of incompetence) (Atherton *et al.*, 2016; and Barthelemy *et al.*, 2022). More specifically, Barthelemy *et al.* (2022) found that about 20% of people with more than one marginalized identity (gender, race/ethnicity, trans) are likely to experience a chilly or hostile climate. Moreover, the authors found that trans physicists experience exclusionary behavior at much higher rates than non-trans physicists. As will be discussed later, this is an area where more research is needed to understand the experiences of graduate students with other marginalized groups, and especially the intersecting effect of these multiple marginalized identities.

In this chapter, we narrow our review to the literature on STEM/physics graduate education by focusing on the main aspects that have been investigated on issues related to equity. We first focus on literature on systemic and cultural factors as barriers to equitable educational experiences. In this section, we primarily focus on institutional actors' approach to equity, diversity, and inclusion, approaches to admission practices, and discuss the state of research in understanding graduate student retention. Finally, we discuss the progress in the literature on understanding the experiences of marginalized students in graduate education, expanding on departmental support in terms of mentoring and social and academic support. We also discuss literature on student work–life balance and mental health. We end this chapter by unpacking the literature on how marginalized STEM/physics graduate students experience microaggression and acts of racism and sexism.

4.2 RESEARCH ON SYSTEMIC AND CULTURAL FACTORS AS BARRIERS TO EQUITABLE EDUCATIONAL EXPERIENCES

In this chapter, we discuss students' experiences focusing on aspects of mentoring, social and academic support, work–life balance and mental health, and experiences of microaggressive acts of racism and sexism. In all these aspects of students' life, it is inevitable to avoid touching on the structural factors that shape students' experiences. As argued by Bensimon (2005), the source of unequal educational outcomes for students from historically excluded racial and ethnic communities is rooted in institutional actors' cognitive frames and reflected in their practices. Institutional actors' (faculty members, administrators, and counselors) cognitive views influence the ways in which they approach student admissions, interact with students and select and implement interventions "targeted" in decreasing educational achievement gaps. Therefore, before expanding on students' experiences, we first unpack the literature that specifically focuses on the systemic and cultural barriers posed in the higher education institutions that shape students' experiences and act as barriers to equitable educational access and outcomes.

4.2.1 Research on cultural and structural context in physics/STEM departments

Culture defines what is done, how it is done, and who is doing it (Gutierrez, 2002). A more formal definition states that "the culture of a group can now be defined as a pattern of shared basic assumptions learned by a group as it solved its problems of external adaptation and internal integration, which has worked well enough to be considered valid, and therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems" (Schein, 2010, p.18). By understanding the physics culture, we can better understand how physics is done and who are (and are not) considered to be physicists.

Among the most dominant and widely accepted values in physics is the notion of meritocracy. People espousing the notion of meritocracy assume that opportunities are distributed based on individual talent, effort, and achievement, ignoring one's inherited social privilege. The meritocratic belief held by institutional actors has implications for who is considered a physicist, who is given the opportunity to be a physicist, and who will participate in shaping future research activities. For example, the meritocratic value is consistent with institutional actors' "fixed mindset" that identifies one's innate talent and posits physics as a "brilliance-required" field (Scherr *et al.*, 2017). Even in departments that strive to prioritize diversity in admissions, physics faculty hold elements of a fixed mindset in their admission practices by emphasizing standardized test scores and undergraduate grades as evidence of high intelligence (Scherr *et al.*, 2017). For example, those physics faculty that applies the Graduate Record Examination (GRE) score as a decisive admission criterion share the belief that higher scores reveal one's intelligence and belonging, virtues that they consider predicting success in the field (Posselt, 2016, 2019). However, these performance metrics are biased against marginalized students (women and Students and Color) (Miller and Stassun, 2014; and Miller *et al.*, 2019). As a result, relying on these metrics disproportionately limits access to physics graduate education for underrepresented students and systematically privileges students from overrepresented backgrounds.

4.2.1.1 Admission practices

Research on physics graduate admission practices suggests that there is a recent tendency to shift away from heavily valuing standardized scores (physics subject GRE) and instead focusing on metrics that paint a more holistic picture of the applicant's background (e.g., recommendation letters) (Potvin *et al.*, 2017). However, many institutions still use cut-off physics GRE values by filtering out the initial pool of applicants (Potvin *et al.*, 2017; and Owens *et al.*, 2020) or list the physics GRE as optional while the general GRE is still required (Owens *et al.*, 2020). Both approaches continue to have the problematic implication of making the final applicant pool less diverse. For example, applying a 700-point cut-off value on quantitative GRE automatically excludes 74% of women applicants as opposed to 27% of men, and about 95% of students from ethnic/racial marginalized communities as opposed to only 18% of white and Asian applicants (Miller and Stassun, 2014). Moreover, the application's physics GRE requirement leads students to narrow down the programs they initially intend to apply to, while the use of the "optional" requirement language makes female students still feel the need to submit their scores. In contrast, male students interpret "optional" as optional (Owens *et al.*, 2020). The GRE requirement is the most prevalent component precluding Bridge students from directly applying to graduate programs. The standardized test requirement does not only act as an anxiety stressor but also is a significant financial burden to many marginalized students (Cochran *et al.*, 2017). Despite the ongoing discussion in academia about intensifying the efforts to make STEM disciplines more diverse, Potvin *et al.* (2020) found that about one-third of selective physics graduate programs still do not consider the applicants' gender and race/ethnicity in the admission decisions. Indeed, programs that have goals of increasing gender and/or ethnic/racial diversity may not be aware of mechanisms for which to do this in their graduate admissions process, particularly when it comes to race/ethnicity (Posselt *et al.*, 2019). One measure for addressing the inequity in admissions has been the implementation of bridge programs (Stassun *et al.*, 2011; and Hodapp and Brown, 2018). The structure of bridge programs varies from one program to another. However, most bridge programs share the goal of preparing students for doctoral programs and/or admitting students to doctoral programs through nontraditional admission practices.

4.2.1.2 Student retention

The prevailing culture in STEM disciplines has traditionally favored norms of success associated with the stereotypical characteristics of white male scientists (competitive, individualistic, and solitary practices), which necessitates spaces that favor effects of marginalization for students who do not share the stereotypical scientist characteristics (Ong *et al.*, 2018). The effects of marginalization are manifested in the average completion rates that differ between underrepresented students and the national average. More specifically, the average completion rate for white students in the physical sciences and mathematics is 52%, whereas it is only 37% for African American students (Sowell *et al.*, 2008). Moreover, a study by Quichocho *et al.* (2019) found that the image of the stereotypical physicist (intelligent white male) precludes women students from identifying themselves as physicists since they do not see themselves as "fitting" into the stereotype.

The literature on STEM doctoral student retention has slowly been developing in recent years (Sachmpazidi, 2021); however, it is still far from providing robust insights on the factors that impact graduate education (Curry and DeBoer, 2020). A systematic literature review of the factors that predict retention for underrepresented students identified three main aspects: personal, social, and institutional (Curry and DeBoer, 2020). For example, among the personal factors, the authors identified internal motivation, identity formation, and perceived support (e.g., family encouragement). The social factors included a sense of belonging, discrimination, mentoring, and work–life balance. Finally, under the umbrella of institutional factors, the authors identified campus and departmental culture, access to role models, and access to networking opportunities. While the authors separated these factors into three categories, the literature has shown how students' personal and social factors are shaped by institutional factors (see Arnold *et al.*, 2020). For example, there is robust evidence in the literature indicating that Women of Color leave STEM fields because of the negative interpersonal relationships (isolation, racism, sexism, and microaggressions) with faculty and peers, which create higher barriers than access to structural resources (e.g., financial aid) (Brown, 2000; Ong *et al.*, 2011; and Ong *et al.*, 2018). As a result, Women of Color create counterspaces (safe social spaces) to find support and enhance feelings of belonging (Ong *et al.*, 2018). Finally, the LGBT Climate in Physics report provided further evidence of a correlation between a hostile climate and the risk of leaving the workplace or school (Atherton, 2016). A significant fraction of study respondents considered leaving, reporting negative climate experiences (adverse climate or observing exclusionary behavior) as the reason for their consideration of leaving the workplace or school.

Research in PER on student retention has argued for the need for systemic change to provide community spaces, mentoring structures, improved student–faculty communication structures and protecting students from faculty with deficit-minded frames on equity-oriented issues (Abdurrahman, 2021; and Sachmpazidi, 2021). Students view remedies for improving retention as lying in the waters of systemic change; however, faculty members view the issue to be related to student under-preparedness and, thus, direct efforts to reduce attrition in the admission processes (Owens *et al.*, 2018; and Sachmpazidi, 2021).

4.3 RESEARCH ON STUDENT EXPERIENCE IN GRADUATE SCHOOL

A plethora of research focuses on the academic experience of marginalized graduate students (women from all racial and ethnic groups and Students of Color). Many of these studies were conducted within the last decade and focused on STEM disciplines. Focusing across multiple disciplines allows researchers to access a larger sample of potential participants. Aggregating across STEM disciplines is reasonable under the assumption that several aspects of STEM students' experiences are homogeneous since STEM programs share many characteristics (program structures and representation of faculty demographics). In the following paragraphs, we primarily draw on the literature in STEM graduate education and note the studies conducted in PER when those are cited. Finally, we end this section by summarizing the state of research in PER along the dimensions of these critical aspects of the graduate student experience.

4.3.1 Microaggressions of racism and/or sexism

The literature in STEM education has long unveiled the explicit and subtle racial and gender discrimination and prejudice in higher education institutions (Harper *et al.*, 2011; and Lee *et al.*, 2020). These forms of discrimination and prejudice have their basis in racial and gender stereotypes—"assumptions about people based on their membership in a particular category, used to justify differential treatment" (Miles *et al.*, 2020, p. 1612). Racial and gender representation (or underrepresentation) of students in STEM is often attributed to stereotypes about intelligence and preparation based on their membership in a particular racial or gender category (McGee *et al.*, 2017). Women and Students of Color interact with faculty members and peers in these institutions on the basis of racial and gender stereotypes that often result in microaggressive acts of racism and/or sexism (McGee *et al.*, 2017; and Miles *et al.*, 2020). These acts of microaggressions cause stress and result in a high emotional, psychological, and physical toll on students (McGee, 2020), as well as contribute to the development of impostor syndrome (Sachmpazidi, 2021; and McGee *et al.*, 2022). McGee *et al.* (2022) introduced the term "racialized imposter phenomenon" to refer to the environmental influences that contribute to the development of the impostor syndrome as a result of structural racism.

Women and Students of Color attending STEM disciplines are observing and experiencing the effects of structural and institutional racism and/or sexism (e.g., higher education institutions dominated by white males in positions of power, unequal pay for similar jobs, unwelcoming institutional climates, and gender/racial/ethnic stereotyping). Women physics and astronomy graduate students largely experience gender microaggressions in the academic trajectory (Barthelemy *et al.*, 2016; and Sachmpazidi, 2021). These microaggressions and, in some cases, hostile sexism take many forms (sexual objectification, sexist language, and assumption of inferiority) and are directed by peers and faculty members. Moreover, while the culture in physics departments continues to be a breeding ground for social injustice, research on the implications of large-scale national interventions, such as the APS BP, brings awareness of these systemic problems. BP students report negative racial experiences in their programs, usually directed from a few individuals (Scherr *et al.*, 2020; and Sachmpazidi, 2021). These microaggressive acts of racism are expressed in the form of comments on BP students being academically inferior that harm students' sense of belonging and challenge their persistence in the field (Sachmpazidi, 2021).

4.3.2 Mentoring, social, and academic support

Mentoring is a process that creates opportunities for students to socialize, advance professionally, enhance self-efficacy, gain access to resources, and network (Griffin *et al.*, 2020). Our higher institutions are built around tenure and advancement policies and practices that prioritize research productivity rather than student interactions. These structures create a culture that is less conducive to mentoring and student demands and/or needs (Brunsma *et al.*, 2017). Scholars in equity work have noted that marginalized students seek to engage with faculty who share similar marginalized identities for two main reasons: (1) students believe that they will benefit from the insights of mentors who have likely experienced similar challenges related to their backgrounds and identities (Baker *et al.*, 2014) and (2) students view this faculty as successful role models in the field (Griffin *et al.*, 2010). However, the students' desire to be mentored by faculty from similar marginalized backgrounds is not always met. The lack of diversity in the faculty body is substantial. In particular, women faculty members and faculty members of color in science and engineering comprise around 39% and 9% of faculty positions, respectively (NSF, 2019). In physics, only about 5.3% of faculty members are African Americans/Blacks or Hispanics/Latinx. Moreover, about half of the African American faculty members are employed at Historically Black Colleges and Universities (HBCUs), where two-thirds of all HBCUs grant bachelors in physics as their highest degree (Ivie *et al.*, 2014). Research suggests that marginalized students do not receive equally good mentorship opportunities compared to white students (Brunsma *et al.*, 2017).

Effective mentoring can help marginalized students' socialization in community practices (Griffin *et al.*, 2020). Community practices include laboratory research, presentations, coursework, and teaching assistantships. As noted, marginalized students develop science identity through participation and knowledge acquisition in perceived critical aspects of community practices. Another essential aspect cited in this study was the notion that a good mentor is one who cares about his or her students, emphasizing the importance of the psychosocial dimension of mentoring. One aspect that should be noted is that the type of financial support graduate students receive is linked to their isolation (Lovitts and Nelson, 2000). Graduate assistantships (research and teaching) offer ways in which students can participate in community practices, for example, sharing office spaces and interacting with peers in teaching preparation sessions, which leads to socialization. However, marginalized physics graduate students are more likely to receive fellowships (a type of financial support linked to isolation) or self-support their education (loans and working outside of campus) than non-marginalized students (Sachmpazidi and Henderson, 2021).

The American Physical Society (APS) developed the Bridge Program (BP) to help enhance racial and ethnic diversity in physics graduate education (American Physical Society—Bridge Program, 2022). The program was designed to offer key departmental support structures, known as key components, to help Bridge students succeed in their graduate education. One of the key components was the constellation mentoring model, where students receive mentoring from multiple sources (academic and research advisors and peers). Some BP departments apply these key components across all graduate students. Research has shown that students in BP departments report being better mentored than students in similar (in terms of size and ranking) non-BP departments (Sachmpazidi and Henderson, 2019). BP students are also socially connected with other students in their program with whom they consider to be like them (in terms of sharing similar backgrounds), which is likely to preclude a sense of social isolation (Scherr *et al.*, 2020). Social and academic support is a critical component influencing student experience (Abdurrahman, 2021; and Sachmpazidi, 2021). Social and academic support is evident, for example, by having people at the department care about students' needs, resources, and support to cope with the graduate program requirements and accessible faculty members. However, women students report experiencing a lack of social and academic support compared with their male counterparts. In particular, this aspect was associated with sexism and gender microaggressions toward women originating mainly from faculty members.

4.3.3 Work–life balance and mental health

Many studies have signaled the need for attention on graduate students' well-being (Evans Bira *et al.*, 2018; Cornwall *et al.*, 2019; and Yusuf *et al.*, 2020). Evans *et al.* (2018) in particular called for a mental health crisis in the graduate student population, reporting that graduate students are six times more likely to experience depression and anxiety than the general population. The authors also found a significant correlation between poor mental health and poor work–life balance. For physics graduate students, students report experiencing a lack of work–life balance connected to the department culture and advisors' mentality (Abdurrahman, 2021; and Sachmpazidi, 2021). Students note that faculty members value students who are willing to work as much time as possible, tying workaholism with the element of being "gifted" (Abdurrahman, 2021). Sachmpazidi (2021) conducted interviews with students and chairs from multiple programs concluding that it is primarily the department culture that pushes students to solely focus on the responsibilities of the graduate student role, neglecting other aspects of one's life. For example, students stated that the department encourages students to work more than 60 h a week under the notion that it is an "investment" for their future; the more one works now, the higher the chances for future professional "success." The heavy workload and the competitive nature of graduate school contributed to developing or worsening student mental health issues (anxiety, depression, and panic attacks).

As discussed earlier, students from marginalized backgrounds are often victims of acts of microaggressions, sexism, and/or racism in STEM disciplines (Ong *et al.*, 2018; and Miles *et al.*, 2020). Microaggressions of sexist or racist remarks in the form of interpersonal interactions or institutional practices result in exacerbation of the sense of not belonging and, in turn, in isolation. These experiences have an impact on students' mental health (Miles *et al.*, 2020). In particular, Arnold *et al.* (2020) examined the impact of academic climate on Women of Color mental health. The authors found that a lack of interpersonal support in the academic setting, difficult interpersonal interactions, gendered, racialized, and cultural encounters were among the factors that negatively contributed to students' mental health issues. While there is a slowly growing body of literature in understanding the factors associated with marginalized students' mental health concerns in STEM graduate education, there is still little we know about the experiences that influence marginalized students' mental health in physics, in particular.

4.4 IMPLICATIONS FOR RESEARCHERS

Science, Technology, Engineering, and Mathematics (STEM) disciplines have been used to express power over marginalized populations in the United States. As STEM strives to become more inclusive, there is a parallel, but small, increase in the representation of marginalized students (Mulvey *et al.*, 2019).

Science communities are still far from achieving social justice and equitable access to education and attainment of equal educational opportunities and outcomes. As discussed in this chapter, women of all races and ethnicities and Students of Color face the consequences of the oppressive structures that have been around for centuries and largely still govern the way institutions of higher education are built and operated. Marginalized students often face microaggressions, racism, and sexism in their daily interactions with institutional actors (e.g., peers and faculty members). They also do not have equal opportunities for mentoring and social and academic support as many of their peers do. For example, marginalized students need to be mentored by faculty role models that share similar backgrounds but since the issue of underrepresentation is further exacerbated in faculty representation, this student desire is often not met. Such student experiences are shaped by institutional structures. As argued in the literature, a system approach is needed to tackle many of the negative experiences and outcomes for marginalized students. One of the critical areas that require more work is related to the retention of marginalized students, largely because (a) there is no systematic data on student retention (and for marginalized students) and (b) most research on student retention focuses across multiple disciplines (STEM and non-STEM). Much of the reviewed literature on the experiences of marginalized students (e.g., mentoring and mental health) aggregates results across several disciplines. Physics is the field with the lowest diversity (Sowell *et al.*, 2008). To better understand both the structural problems in the physics graduate programs and how these affect student experiences, there is a need for discipline-based focus qualitative studies under the lens of critical race and feminist standpoint theories (Rodriguez *et al.*, 2022).

In this chapter, we primarily focused on the experience of marginalized students with certain identities (gender, race/ethnicity). However, students with disabilities and LGBTQ+ students experience the results of similar institutional structures that preclude their participation in physics. As noted early in the Introduction, little is known about the experiences of these students, especially in graduate education. There is a significant gap in the literature that needs to be addressed in order to mitigate the systemic inequities in physics.

ACKNOWLEDGMENTS

The author would like to thank Dr. Geraldine L. Cochran for her detailed feedback and Dr. Chandra Turpen for her encouragement and support during the authorship of this chapter. The author acknowledges financial resources as part of the Effective Practices for Physics Programs (EP3) Initiative from NSF No. 1821372 and the APS Innovation Fund that supported her during the time of authoring the chapter.

REFERENCES

Abdurrahman, F. N., Ph.D. dissertation (University of California, Berkeley, 2021). See https://escholarship.org/uc/item/1zs693tt
Arnold, A. C. *et al.*, Proceedings of the 2020 American Society for Engineering Education Conference (ASEE, 2020).
Atherton, T. J. *et al.*, *LGBT Climate in Physics: Building an Inclusive Community* (American Physical Society, College Park, MD, 2016).
Baker, V. *et al.*, Int. J. Res. Dev. **5**(2), 83–98 (2014).
Barthelemy, R. *et al.*, Phys. Rev. Phys. Educ. Res. **12**, 020119 (2016).

Barthelemy, R. *et al.*, Phys. Rev. Phys. Edu. Res. **18**, 010124 (2022).

Bensimon, E. M., New Dir. Higher Educ. **2005**, 99–111 (2005).

Brown, S. V., *Access Denied: Race, Ethnicity, and the Scientific Enterprise*, edited by G. Campbell *et al.* (Oxford University Press, New York, 2000).

Brunsma, D. L. *et al.*, Sociol. Race Ethnicity **3**(1), 1–13 (2017).

Cochran, G. *et al.*, paper presented at *Physics Education Research Conference, Cincinnati, OH* (AAPT, 2017). See https://www.compadre.org/Repository/document/ServeFile.cfm?ID=14576&DocID=4753.

Cornwall, J. *et al.*, Stud. Continuing Educ. **41**(3), 363–380 (2019).

Curry, N. F. and DeBoer, J., paper presented at *2020 ASEE Virtual Annual Conference Content Access, Virtual Online* (ASEE, 2020).

Evans, T. M. *et al.*, Nat. Biotechnol. **36**(3), 282–284 (2018).

Ghee, M. *et al.*, CBE Life Sci. Educ. **15**(3), ar28 (2016).

Griffin, K. A. *et al.*, *Socialization in Higher Education and the Early Career. Knowledge Studies in Higher Education*, edited by J. Weidman and L. DeAngelo (Springer, Cham, 2020), Vol. 7.

Griffin, K. A. *et al.*, New Direct. Inst. Res. **2010**, 95–103 (2010).

Gutiérrez, K. D., Human Dev. **45**(4), 312–321 (2002).

Harper, S. R. *et al.*, J. Coll. Student Dev. **52**(2), 180–200 (2011).

Hodapp, T. and Brown, E., Nature **557**, 629–632 (2018).

Ivie, R. *et al.*, Technical Report, American Institute of Physics, College Park, MD, 2014.

Lee, M. J. *et al.*, Int. J. STEM Educ. **7**, 48 (2020).

Lovitts, B. and Nelson, C., Academe **86**(6), 44–50 (2000).

McGee, E. O., *Black, Brown, Bruised: How Racialized STEM Education Stifles Innovation* (Harvard Education Press, Cambridge, MA, 2020).

McGee, E. O. *et al.*, J. Divers. Higher Educ. **10**(3), 253–270 (2017).

McGee, E. O. *et al.*, Race Ethnicity Educ. **25**(4), 487–507 (2022).

Miles, M. L. *et al.*, J. Res. Sci. Teach. **57**, 1608–1631 (2020).

Miller, C. and Stassun, K., Nature **510**, 303–304 (2014).

Miller, C. W. *et al.*, Sci. Adv. **5**(1), eaat7550 (2019).

Mulvey, J. *et al.*, Technical Report, American Institute of Physics, College Park, MD, 2019.

National Science Foundation, National Center for Science and Engineering Statistics, Special Report NSF 17-310, Arlington, VA (2019). See https://ncses.nsf.gov/pubs/nsf21321/report/academic-careers

Oakes, J., Rev. Res. Educ. **16**(1), 153–222 (1990).

Ong, M. *et al.*, J. Res. Sci. Teach. **55**, 206–245 (2018).

Ong, M. *et al.*, Harvard Educ. Rev. **81**(2), 172–209 (2011).

Owens, L. *et al.*, *2018 PERC Proceedings*, Washington, DC, 1–2 August 2018 (AAPT, 2018).

Owens, L. *et al.*, paper presented at *Physics Education Research Conference 2020, Virtual Conference,* 22–23 July 2020 (2020). See https://www.compadre.org/Repository/document/ServeFile.cfm?ID=15513&DocID=5361 (last accessed May 25, 2021).

Posselt, J. R., *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping* (Harvard University Press, 2016).

Posselt, J. R. *et al.*, J. Women Minor. Sci. Eng. **25**(4), 283–306 (2019).

Potvin, G. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 020142 (2017).

Quichocho, X. R. *et al.*, *2019 PERC Proceedings, Provo, UT, July 24-25, 2019*, edited by Y. Cao *et al.* (AAPT, 2019).

Rodriguez, M. *et al.*, Phys. Rev. Phys. Educ. Res. **22**(1), 013101 (2022).

Rosa, K. and Mensah, F. M., Phys. Rev. Phys. Educ. Res. **12**(2), 020113 (2016).

Sabella, M. S. *et al.*, Phys. Teach. **55**, 350–355 (2017).

Sachmpazidi, D., Ph.D. dissertation (Western Michigan University, 2021). See https://scholarworks.wmich.edu/dissertations/3771

Sachmpazidi, D. and Henderson, C., paper presented at *Physics Education Research Conference 2019, Provo, UT* (AAPT, 2019). See https://www.compadre.org/Repository/document/ServeFile.cfm?ID=15333&DocID=5246 (last accessed May 25, 2021).

Schein, E. H., *Organizational Culture and Leadership*, 4th ed. (Jossey-Bass, San Francisco, CA, 2010).

Scherr, R. E. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020132 (2020).

Scherr, R. E. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 020133 (2017).

See https://www.aps.org/programs/minorities/bridge/index.cfm for (last accessed August 23, 2022).

Smith, K. C. *et al.*, J. Negro Educ. **88**(3), 407–418 (2019).

Sowell, R. S. *et al.*, *Ph.D. Completion and Attrition: Analysis of Baseline Demographic Data From the Ph.D. Completion Project* (Council of Graduate Schools, 2008).

Stassun, K. G. *et al.*, Am. J. Phys. **79**(4), 374–379 (2011).

United States Census Bureau, see https://www.census.gov/quickfacts/fact/table/US/SEX255221#SEX255221 under subheading "Education" for women population with a "Bachelor's Degree or higher."

Yosso, T. J., *Critical Race Theory in Education* (Routledge, 2016), pp. 113–136.

Yusuf, J. *et al.*, J. Public Affairs Educ. **26**(4), 458–483 (2020).

CHAPTER 5

RESEARCH DESIGN CONCERNING EQUITY IN PHYSICS EDUCATION RESEARCH (PER)

Alexis V. Knaub and Lin Ding

Knaub, A. V. and Ding, L., “Research design concerning equity in physics education research (PER),” in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 5-1–5-28.

5.1 INTRODUCTION

5.1.1 What does equity in research design have to do with physics education research (PER)?

This chapter is focused on *equity in research design for PER scholarship*. We focused on *equity*, rather than equity, diversity, and inclusion (EDI).[1] Although inclusion and diversity can be part of equity and are often used interchangeably, as pointed out by Cochran (2018), they are not the same. That said, equity does not have a universal definition. As indicated in a study by Rodriguez *et al.* (2012), equity can have multiple definitions and lead to different results depending on the definition. Based on multiple readings (e.g., Rodriguez *et al.*, 2012, Cochran, 2018; CSSP, 2019; and The Anne Casey Foundation 2022), we defined equity as focusing on dismantling systemic issues, justice, and ensuring that people have what they need to succeed.

Research design sets the direction and course of a study from start to finish to answer research questions. Researchers consider methodological choices (e.g., qualitative, quantitative, or mix/multiple methods), sampling, data collection, and analysis techniques when creating a plan to best answer the research

---

[1] We are aware of other variations of the EDI acronym, including newer variations of the acronym that have “A” for accessibility, “B” for belonging, “J” for justice, and “SJ” for social justice. Although these ideas are important and were in our thinking for this work, we do not wish to focus this chapter on the evolution of the EDI acronym and its implications.

https://doi.org/10.1063/9780735425514_005

questions of interest. We have included researchers and research teams in this definition as well. Who is on the research team can impact what research areas are studied (Gerwin, 2018).

For the purposes of this chapter, we define *equity in research design for PER* as dismantling systemic issues that create and sustain barriers, hindrances, or harm for people with marginalized identities during the creation and planning of a PER study. People with marginalized identities are those whose power is systemically limited. We consider two areas that research design impacts equity. One is those directly involved with the study: researchers and research participants. The other is the research's impact on equity. PER studies impact people's experiences in physics spaces, from individual educators' practices to national-level policy. Designing studies that have equity in the foundation and weave equity considerations throughout the study can support researchers in making more equitable choices with more equitable outcomes.

These base-level decisions impact all PER, not just studies that are considered EDI work. For example, Kanim and Cid (2020) examined published work in PER and found evidence suggesting that the participants in published PER tend to be white and affluent undergraduates at four-year institutions, and have considerable mathematical preparation. This work indicates that the PER manuscripts are not representative of most physics students, calling into question whether the findings are generalizable to all students. Without more diverse populations and contexts, as well as careful studies to understand possible reasons for results, PER could inadvertently advocate for practices that are ineffective or even harmful. In summary, regardless of whether a study is focused on EDI, the importance of equity is clear.

5.1.2 Writing process for this chapter

We drew upon the literature from PER and relevant education and social science research. PER is a relatively new field at the time of writing. Other fields have more mature bodies of literature around topics that we either know are relevant to PER or are likely to be so. Much of our literature search focused on marginalized populations. This includes race; ethnicity; gender; socioeconomic class; Lesbian, Gay, Bisexual, Trans, Queer, and related communities (LGBTQ+); countries of origin; and accessibility/disability. We also broadly read the literature on vulnerable populations. When relevant and specified in the work, we named the specific population of the work cited. Some practices are applicable to multiple populations, hence not being more specific.

Knaub did the majority of defining the scope of this chapter, the literature search, and chapter writing. Ding contributed to each of these areas.

Although this chapter is not a traditional manuscript, we would be remiss not to include our own positionalities as those informed the chapter.

- Knaub is an East Asian (South Korean) woman who was adopted at 10 weeks of age and grew up in a white working-class household; these identities give her some firsthand experiences regarding marginalization and complexity around identity. EDI is integrated into her research, external evaluation work, and service endeavors. This combination of the personal and professional experiences provides her with many experiences that inform her perspectives and knowledge around EDI work, with beliefs that equity is integral to everything and that much of the needed equity work is non-linear and not necessarily visible.

- Lin is a Chinese man. He was born and raised in Shanghai, the biggest metropolitan area in China. He came to the United States in his early 20s to pursue higher education and since then has lived and worked in the U.S. as a first-generation immigrant resident. Because of his cross-cultural identity, he is at a unique vantage point to witness and experience complex power dynamics, implicit bias and microaggressions in the U.S. academia. This has shaped his understanding of EDI and led him increasingly to believe in the importance of having his voice heard.

5.1.3 Sections in this chapter

To cover equity in research design for PER studies, we cover multiple facets of research design processes and different stakeholders as well as the ways that they may impact equitable research design. Each section has some reflection questions for researchers to consider. These questions do not have a "right" answer but are designed to support researchers in making more informed decisions.

We begin with discussing researchers and the formation of research teams because *who* does the research impacts all facets of the research from what research questions are asked and how data are collected to what analysis occurs. This includes not only the positionality of the researcher or research team but also the complex power dynamics that may exist for the researchers and the research participants (e.g., the researcher may be the professor of potential research participants). Research questions are the next topic, looking at areas such as what research questions are asked and consideration for underlying assumptions. The ways in which research questions are posed may introduce inequities (e.g., research questions that suggest marginalized populations are deficient compared to dominant populations).

Methodological choices are discussed, both general considerations (e.g., sampling bias) and specifically around quantitative work (e.g., for whom instruments are valid, reliable, and fair) and qualitative work (e.g., interview techniques that might be uncomfortable for marginalized people). Matters around research participants and samples are also discussed. The data collection section focuses on both the researchers and the research participants. Both researchers and the research participants can be impacted (e.g., interview studies where the researcher and research participants share a marginalized identity or traumatic experience might be triggering for both the researcher and researched). This section also discusses what data are collected (e.g., racial demographic data) and how they are collected (e.g., passive data collection where the people being researched may not be aware their data are being used). We also address matters of safety for both researchers and research participants.

The analysis and results sections discuss how the analysis is conducted and how results are presented. For analysis, we focus on potential issues around analysis choices such as collapsing categories (e.g., categorizing Asians with white students) and how positionality can impact analysis interpretation (e.g., a researcher might not accurately interpret observed behaviors from Deaf research participants if the researcher is not Deaf). The section on research results considers how results are discussed and interpreted (e.g., what limitations to generalizability are mentioned). The penultimate section is about potential changes to research design. Because research projects are dynamic, there may be changes needed. Such changes may introduce inequities (e.g., a faster timeline might put strains on researchers). Another area addressed by this section is actions and considerations that researchers might take if the research is found to be harmful and inequitable.

This chapter concludes with final considerations, reflection, and learning around equity. Equity in research design is not easy or straightforward yet needs to be a focus in order to mitigate harm from PER studies and to maintain the highest level of rigor in scholarly work.

The format of this chapter may suggest neatly defined categories and a linear process. We wrote in this fashion for readability and anticipate that readers might find each section useful at different points in their research process.

5.2 MAKING THE TEAM: WHAT TO CONSIDER WHEN INITIATING A STUDY ON EQUITY

The importance of the researcher and the researcher team could not be overstated. The researcher or team has considerable power in making all sorts of decisions from the very beginning. They steer the direction and course of the work regarding what to study, which methodology to use, what data to collect (and what are considered data), how to analyze and interpret data, and when, where, and how to share findings (Mertens and Ginsberg, 2008; D'Ambrosio *et al.*, 2013; and Parson, 2019).

Positive pre-existing relationships between the researchers and researched can make access to different populations or settings easier (Quinn, 2015). Though largely beneficial if the relationships are positive, pre-existing relationships add complexity. The reasons that positive pre-existing relationships are an important range from gaining permission to conduct research in a setting where one does not work to having established trust already; lacking pre-existing relationships may mean the research might not be possible. When conducting research on colleagues and peers, there can be clear advantages such as familiarity but also disadvantages such that prospective participants may feel coerced into participating (McDermid *et al.*, 2014). We can imagine this complexity being even messier for PER studies conducted at one's institution, department, or class where the researcher is also involved with the participants' professional lives (e.g., the researcher is the department chair, teaches the class being studied, etc.) This makes them indeed part of the research design, and thus decisions around who is conducting the research are foundational. If the researchers and the researched had unpleasant past relationships, addressing the issues in the early stages is prudent. Because such issues are idiosyncratic and nuanced, seeking advice and support from experts and trusted colleagues will likely help future research activities.

The decisions made by researchers and research teams are impacted by their identities, because identities shape experiences and perspectives. Our identities can lead to more equitable or inequitable research by impacting what research is done, how it is conducted, whether individuals choose to participate, and to what extent research and recommendations are useful for the population being researched. The interviewer's behaviors and identities (e.g., gender) can impact what is shared in an interview (Pezalla *et al.* 2012). Shared identities can be important even if the researchers do not have pre-existing relationships with the research participants. For data collection, working with researchers who share relevant marginalized identities with the participants can help create rapport and otherwise missing trust if the researchers do not share the exact same identities as the participants (Wilson and Neville, 2009).

5.2.1 Understanding the researcher and the research team's potential impact on the research process

Knowing specifically how the identities of researchers or a research team might impact the research may not be readily obvious. Writing a positionality statement at the beginning stages can be useful for researchers. Positionality statements are sometimes included in PER manuscripts to help the audience understand how the researchers' identities, experiences, and expertise shaped the research. Though these are more common in qualitative work, researchers have been advocating that they be included in quantitative studies because researcher perspective impacts multiple parts of quantitative work including the collected variables and the interpretation of results (e.g., Pearson *et al.*, 2022).

Considering positionality at the beginning can support the researcher or team in making informed choices for the design. This suggestion is not meant to create a checkbox exercise or imply that one individual/team is analogous to an entire demographic, but it is intended for researchers and research teams to meaningfully engage with how their various identities create opportunities, challenges, or otherwise impact the research. Zhang and Du (2021) noted that when conducting intersectionality studies, people have many identities, so researchers should consider which identities are relevant to research participants. Evans-Winters and Esposito (2019) advocate for critical reflexivity when using an intersectional perspective, asking "What is the researcher's social political proximity to the educational problem or issue, and to the research participants themselves?" Although the context for the work cited here is specific [i.e., Zhang and Du's (2021) work on intersectionality at the data collection stage and Evans-Winters and Esposito's (2019) qualitative work on intersectionality], their recommendations are relevant to all beginning stages of research design including consideration of the research team.

Care should be taken to consider specific identities and the extent to which those identities are truly shared with the researched population of interest. For example, while different identities are included in LGBTQ+, the individual identities in the acronym have some unique experiences (Griffith *et al.*, 2017). Similarly, the various identities that comprise other aggregate categories (e.g., People of Color, people with disabilities and unmet accessibility needs) are not interchangeable. The differences and uniqueness are amplified when considering intersections of other identities. For example, the authors of this chapter are both East Asian but differ in gender, culture, language, nationality, and immigration path.

5.2.2 Inviting additional researchers to the team

Examining positionality may help the researcher or team realize that the study would be problematic if it continues with the current team. One researcher does not need to have all relevant knowledge, experience, and skills to avoid pitfalls around research design. In the National Academies' Discipline-Based Education Research (DBER), the authors suggest that working with others can be a productive step:

> High-quality DBER combines expert knowledge of a science or engineering discipline, of the challenges of learning and teaching in that discipline, and of the science of learning and teaching generally. This expertise can, but need not, reside in a single DBER scholar; it also can be strategically distributed across multidisciplinary, collaborative team (p. 2). (National Research Council, 2012)

This sentiment has also been voiced specifically in physics education. Boveda and Weinberg (2020) pointed out that scholars in disciplines such as education, gender studies, and other disciplines outside physics education can provide important perspectives in curriculum and pedagogy and bring in valuable expertise. Collaborating with such scholars can be useful for areas beyond curriculum and pedagogy, as scholars in PER study a multitude of areas within the physics context, such as organizational change and experiences of marginalized individuals.

While there are ways to include missing perspectives from marginalized people such as creating advisory boards (Griffith *et al.*, 2017) or having listening sessions with the population being researched (Gaddy and Scott, 2020), including someone on the research team who has relevant identities and expertise could be a productive step to address some equity concerns. Parson (2019) points out "In each step of the research process, researcher decisions take agency, and therefore power away from participants." Researchers who share identities, experience, and expertise with those of the participants are not a substitute for an advisory board, a listening session, or other means of including feedback from participants. However, given the numerous decisions that occur during the research process, it may not be viable to involve participants in every decision.

Research participants often do not experience benefits equivalent to those of the researcher (Parson, 2019). For example, researchers can gain career advancements and accolades because they have published studies. This can accumulate in subsequent opportunities and awards (e.g., publications can strengthen grant applications) that can set the researcher's entire career on a course for success. Research participants will likely not reap such long-term benefits. For example, students enrolled in an introductory physics course may not directly benefit from any suggestions they make for subsequent courses. Balancing input and feedback from participants with constraints is important to ensure that participants are not overburdened.

One more equitable option is to establish collaborations with researchers who both have expertise in the research area of interest and have similar or the same marginalized identities with those of the research population of interest. Research partnerships can provide publication opportunities, which are valuable in academia for career advancement. Such partnerships can also support researchers in expanding their professional networks, which can lead to more opportunities for collaborations and invited speaking engagements. While this may not provide benefits to the research participants, such collaborations can be beneficial to the population of interest. It can be meaningful for people with marginalized identities to see someone with shared marginalized identities succeed. A researcher with marginalized identities may be suited to advocate for people with similar identities and remove structural challenges that create marginalization. Although individuals in less privileged positions do work to remove structural challenges, having established credibility and professional networks are an asset in changing systems.

Simply adding researchers with relevant experience, expertise, and identities is not adequate if one of the goals is to be more equitable. A research collaboration cannot be equitable if the research environment is toxic. As revealed from the report by the National Academies (2018), The Science, Technology, Engineering, Mathematics, and Medicine (STEMM) disciplines have pervasive sexual harassment issues in research spaces, among other STEMM spaces (National Academies, 2018). Although the report did not specifically focus on PER, there is no reason to believe that these do not occur in PER, along with other oppressive behaviors.

Even if harassment is not present in the research environment, additional researchers do not mean that the research will be equitable. The research partnership should be a collaboration where the expertise is included and valued, as it would be with any collaboration. The collaboration may also need to work through team dynamics. Teams that have more diversity can be more creative and innovative than teams of homogeneous members, but can also experience more conflicts (Plowman *et al.*, 2007; and Stahl *et al.*, 2010). However, conflict is not necessarily bad if it is well-managed (Uhl-Bien *et al.*, 2007). Being cognizant of power dynamics that can arise and being proactive in handling conflict can support the team in being equitable.

Deciding upon a research team at the beginning of the work does not mean that reflecting on ourselves as researchers is completed. Re-examining positionality throughout the research process can help researchers understand their impacts on the research participants and on the population of interest so that they do not reify inequities (Parson, 2019). Depending on the duration of a study, a researcher's positionality can change. For example, researchers may experience a change in socioeconomic status by being hired in a well-paying job, which complicates their identities.

5.2.3 Addressing current and historic inequities

No research occurs in a vacuum. Researchers before us conducted PER studies, and individuals engaged with researchers in PER in many facets, such as data collection or reading published work. Pearson *et al.* (2022) stated

> STEM equity researchers must grapple with the historical events that have shaped what STEM fields look like today. STEM environments have led to scientific discoveries and innovation, but these environments also have a history of reproducing systemic inequities that harm individuals.

Rodriguez *et al.* (2022) make a similar statement:

> More specifically, a scientifically literate population is critically cognizant of how science has been (and is being) used and abused to oppress marginalized populations, including communities of color, especially those living in or near poverty.

We modify and expand on these arguments for the consideration of researchers in PER: researchers in PER must grapple with current and historic interactions and work in physics education that can (and do) oppress marginalized populations. Researchers in PER might not see themselves as equity researchers, but as suggested by Kanim and Cid (2020), their work impacts marginalized students. Educators can and do implement practices in physics learning spaces with marginalized students. The impact is not confined to educators implementing practices but interactions such as researchers engaging with students in classrooms or physics educators at conferences.

Current and historic interactions and work in PER impact people, who will in turn impact current and future research. Grappling with the impact of current and past inequities stemming from PER is nuanced and complex. Researchers should consider how they individually and collectively have impacted marginalized populations through various interactions and work. These considerations include not only the individuals who are a part of the research team but also the broader impact of researchers in PER. Researchers may address the current and historic impacts of PER by employing some of the strategies described earlier in this section.

Questions for reflection

1. Who is on the research team and why?
2. What is the positionality of each researcher? How will the positionality of the researcher impact each aspect of the research?
3. What are some critiques or challenges researchers have faced in this research area? How do the positionalities of the researchers for the project address the critiques and challenges?
4. Are there important perspectives missing from the research team?
5. How will the research team involve each member in a meaningful manner?

5.3 DEVELOPING RESEARCH GOALS AND QUESTIONS THAT BENEFIT THE POPULATION BEING STUDIED

Research goals and questions define the scope of the study and what might be learned. The framing of these goals and questions can support methodological choices, including what data are collected.

Selecting what to study and consequentially what not to study prioritizes certain aspects of a topic. Determining what to prioritize may introduce inequities. The research questions may not be of interest or beneficial to the population being studied (Griffith *et al.*, 2017). Research questions may be posed in a way that suggests that the population being researched needs to assimilate into a dominant culture or contribute to othering a population (Griffith *et al.*, 2017). Research focusing on whether marginalized people are meeting the same metrics as people from dominant cultures can be problematic (Rodriguez *et al.*, 2012; San Pedro and Kinloch, 2017; and San Pedro, 2021). This is often referred to as "deficit research" or "gap research," which assumes that there is a universally good standard that the dominant group has achieved and that marginalized groups should meet too. Deficit research has been critiqued in PER for these reasons (e.g., Rodriguez *et al.*, 2012; and Traxler *et al.*, 2016). Such research ignores other possibilities that are indicative of a desired result. For example, San Pedro (2021) points out that Indigenous parents are involved in their children's lives but not in ways that schools would identify. Thus, researchers could make incorrect assumptions regarding parental involvement if they are not careful.

There are ways to generate more equitable research questions that a population is interested in. Involving community members of the population of interest from the beginning and throughout the research process can ensure that the study is of interest to the larger population (Merten, 2010; and Gaddy and Scott, 2020). Researchers in PER may have unarticulated assumptions, hypotheses, or theories guiding their work (Knaub *et al.*, 2019). Articulating these can help researchers consider what they are asking and any potential limitations to the questions. For example, researchers are often interested in understanding whether students are learning physics better after implementing a particular technique in a course. Clarifying what many of these terms mean would be beneficial to explore unarticulated assumptions and bolster support for any hypotheses. Researchers should consider not only how they are measuring learning (e.g., doing well on an exam, being able to use knowledge in a different context) but also what they assume in the research goals and questions. Learning physics can mean many things (e.g., problem solving, cultural aspects of physics, or something entirely different). These choices can support equity by more expansive definitions or reify inequities through definitions that unnecessarily exclude what learning physics means and how we measure learning physics.

Questions for reflection

1. Why are the research goals and questions important to answer?
    a. For whom are these goals and questions important?
    b. Under what circumstances are these goals and questions important?
2. What assumptions are built into these questions and goals?
3. How might the research goals and questions, as they are phrased, reinforce inequity?
    a. What possible inequities (or biases) may be introduced in the goals and questions?
    b. Are there alternative ways to phrase a research question that is more equitable?

5.4 RECRUITING AND SELECTING RESEARCH PARTICIPANTS, PARTICULARLY FROM MARGINALIZED POPULATIONS

Research participants provide the data; without them, there would be no studies. Yet, they may not truly benefit from the studies in which they participate. As noted above, research participants often do not experience equivalent benefits to the researcher (Parson, 2019). In the PER context, recommendations may be implemented long after the participants are enrolled in a physics course or program. We focused on being *more* equitable, recognizing that it may not be possible to be completely free of inequities in PER studies. This section focuses on what researchers should consider when selecting and recruiting study participants.

5.4.1 Conceptualizing the research sample and recognizing potential challenges in recruiting research participants

As discussed earlier in this chapter, not having participants from marginalized populations raises questions such as for whom the research is applicable and can produce or exacerbate inequities for marginalized populations. Researchers should strive to include a wide range of individual research participants, as there are multiple perspectives from people with shared identities (Griffiths *et al.*, 2017). Important information is missing when intersectional identities are not included in the research (Griffiths *et al.*, 2017; and National Academies, 2018). For example, the experiences of Women of Color are different from white women because Women of Color can experience oppression from the interplay of race and gender. A large sample is not indicative of a diverse sample (Griffiths *et al.*, 2017).

Researchers should be careful that the research sample can truly answer the research questions or represent the population of interest. Because this is a nuanced issue that is largely determined by the research questions or goals, there should not be a one-size-fit-all type of straightforward guidance about recruiting representative samples. For some studies, it is completely reasonable or even necessary to recruit research participants who are not representative of physicists at large (e.g., studies that focus on people with marginalized identities). For studies that are not designed to focus on one population, researchers should consider where and how they are seeking research participants and what potential limitations may be of where they are recruiting participants. Revising the research questions and goals to be more specific is a possible way of ensuring that the findings are not accidentally generalized to all people.

Ensuring that the research sample includes participants with marginalized identities is not simply a matter of advertising to them or sending an invitation. While participation of individuals with marginalized identities can be important and beneficial to the studied population, the research study, and the field, researchers should know that individuals with marginalized identities are not obligated to participate (Parson, 2019). There are many reasons that an individual, especially one with marginalized identities, may not participate because of not having the time or not being interested in the study. Seeking out already-collected data that could answer the research questions is a way to be less burdensome to individuals (Gaddy and Scott, 2020).

Building or having positive relationships with communities of interest is important for research. This theme has appeared in research involving Indigenous people (Ndimande, 2012; Muhammad *et al.*, 2015; and San Pedro, 2020) and generally when researchers would be considered outsiders (i.e., do not share relevant identities or experiences with the population of interest) (Joseph *et al.*, 2021). Given the potential impacts of the research on the immediate participants and future applications of the research, it is understandable that prospective research participants with marginalized identities may be cautious. Prospective research participants may be asked to share sensitive information and be concerned that the research could be used to harm people in their communities, regardless of the researcher's intentions. If the researchers have no relationship with the prospective participants, prospective participants may choose not to participate because they may not inherently trust the researchers to mitigate these issues; therefore, as discussed earlier in this chapter, the research team and their relationships to communities of interest matter.

While researchers advocate that individuals from marginalized research participants are included in research and intentionally invited, researchers need to be cautious regarding the use of typical research sampling methods. One example is snowball sampling, which can accidentally reveal LGBTQ + identities to others (Griffiths *et al.*, 2017). Revealing information can also impact people who are undocumented, victims or survivors of various types of harassment, anything that might be embarrassing to someone (e.g., failing a course), or other situations that expose the vulnerability of prospective participants. Care should be taken so that other prospective participants do not see who else is included in invitations. Researchers on teams should consider which team members have access to raw or identifying data. They should also be transparent in informing prospective participants who will see the identifying data.

5.4.2 How the participants can be decision-making partners in studies

In the section on researchers and research teams, we advocated collaborations with researchers who have similar marginalized identities to those being researched. They may have similar experiences to that of the researched, which can help identify potential problems. However, these collaborations may not be adequate. Even if researchers have similar marginalized identities to that of the participants, they can still marginalize the participants (Parson, 2019). Researchers may have privileged identities that the research participants do not, which may make researchers less aware of potential issues with how they conduct or present research. Secules *et al.* (2021) pointed out that even with shared marginalized identities and empathy for research participants, researchers are still positioned differently than that of the research participants and can be more privileged in this context. For example, researchers may be faculty members and the research participants may be enrolled in the faculty member's course

For these reasons and more, various individuals advocate for going beyond the researcher-participant binary by working collaboratively with research participants (e.g., Griffiths *et al.*, 2017; San Pedro and Kinloch, 2017; San Pedro 2020; and Dounas-Frazer *et al.*, 2021). Some of these methods are part of decolonizing (e.g., Ndimande, 2012; and Sandoval *et al.*, 2016) or indigenizing research (e.g., San Pedro, 2020) methodologies that focus on community relationships, blurring the demarcation between researchers and research participants. There are different ways of varying involvement where research participants can have input on the research. When possible, research participants should be able to self-identify (Griffiths *et al.*, 2017). Seeking consent and feedback on how research participants are described and portrayed in publication gives agency to participants, who may have reasons for why they prefer certain phrasing or descriptions be included or omitted (e.g., Griffiths *et al.*, 2017; San Pedro and Kinloch, 2017; and San Pedro 2020). This issue may be especially important for qualitative research, where rich stories are portrayed, but it can be an issue for quantitative research to determine which details are included, for example, descriptions of samples or settings. The description can perpetuate stereotypes or be inaccurate from the participant perspective.

Although researchers should be mindful of what they are asking of research participants and how their time is being used, marginalized populations are interested in being involved in the research process. Thus, researchers should not assume that participants will not want to participate in facets of the research process but should consider ways that are meaningful to seek their involvement. Some ways to include research participants are feedback on research findings from participants (Dounas-Frazer *et al.*, 2021); co-constructing the research narratives with the participants (San Pedro, 2021); partnering with the participants in ways such as research-practice partnerships (Dounas-Frazer *et al.*, 2021) or students-as-partners in research (Sohr *et al.*, 2020).

Involving research participants in the research is complex. Although the above strategies are intended for the research participants to have more agency in the research, the reality can introduce difficulties. Sohr *et al.* (2020) discussed times where even with the best intentions, decision-making opportunities may not be equitable for research participants; they noted that they need to repair relationships at times, suggesting that some situations can be rectified. Researchers may need to develop new collaboration skills to ensure that individuals from different backgrounds are truly in an equitable collaboration (Chapman and Ainscow, 2019).

Although this may seem more of an issue during the data collection, data analysis, or publication stages of studies, prospective research participants may want to know from the beginning whether they will be able to provide input on the study. Planning and communicating how research participants can be involved in decision making around the study gives more agency to the participant and may increase the likelihood of their participation.

Questions for reflection

1. Who is being invited to the study?
2. What is the relationship between the participants and the research team?
3. To what extent do the research participants reflect the population of interest?
4. In what ways can the participants give feedback and make decisions on the research?
    a. What are the different points at which participants can give feedback?
    b. How are researchers ensuring that these feedback mechanisms and decisions making structures diffuse or flatten hierarchical structures?

5.5 RESEARCH INSTRUMENTS

Research instruments provide a means for collecting data. In PER, researchers develop their own instruments or use instruments that were developed by other PER or social science scholars.

5.5.1 Considerations when designing instruments

As mentioned previously, research participants may not directly benefit from the study. Researchers should be mindful of the questions they are asking and whether they need such data, especially if the questions may involve considerable mental or emotional efforts to answer (Gaddy and Scott, 2020). Questions that delve into a research participant's experiences may be uncomfortable as they may elicit vulnerable or traumatic experiences. One strategy aimed at providing some comfort to participants with marginalized identities is to ask broad questions about the population of interest rather than about their specific experiences. This can give participants an option to share their own experiences or observations that they have had (Zhang and Du, 2021). For example, a researcher in PER might ask "What are your impressions of this department for international students?"

5.5.2 How to include demographic questions on instruments

Studies often collect demographic data on the research participants to gain some insight into who participated. Using current demographic terminology that is considered respectful by the population being described is important to ensure that research participants are respected (Griffiths *et al.*, 2017). If the instrument contains demographic questions, such questions may not always be valid across multiple cultural contexts, particularly international contexts because people have different terminologies (Griffith *et al.*, 2017). Anteneodo *et al.* (2020) studied the landscape of Brazilian physicists and drew upon the Brazilian Institute of Geography and Statistics census terminology for race, which uses skin color as a means for identifying race.

Furthermore, researchers advocate that disaggregated data across broad demographic categories (e.g., providing options such as Chinese, Korean, Vietnamese, etc., and not just "Asian") can provide important insights into equity (Museus and Kiang, 2009; Griffiths *et al.*, 2017; and Teranishi and Kim, 2017). Disaggregation is important for intersectional data as well because of the unique experience individuals at various intersections of identity experience. However, collecting these data may make research participants concerned about being exposed as participants if they have marginalized identities and few individuals share these identities, as is the case in physics. Census style studies in physics have shown that there are few African American, Hispanic American, and/or women at various levels of physics (AIP, 2018). The small number means that individuals are more likely to be identified. Our point is not to prevent researchers from collecting these data or disaggregating categories. Rather, we suggest exercising caution and being aware of these issues to make good decisions.

Some researchers go further with inclusive demographic categories and suggest that research participants self-define rather than select from pre-defined demographic categories (Griffiths *et al.*, 2017; and Strunk and Hoover, 2019). This gives considerable agency to research participants, but it may be difficult to describe the collective group of research participants if there are few commonalities among definitions. This may also further burden the research participants to write or state their identities rather than simply pick from a few options. We are not advocating against self-defining, just pointing out that it could introduce other issues for the research participants.

5.5.3 Understanding equity issues that are present for research instruments aimed at large-scale assessment

There are many research tools in PER ranging from concept inventories [e.g., the Force Concept Inventory (FCI)] to observation and interview protocols. The availability of these instruments can lower the barrier of conducting research, as it takes considerable time and expertise to create a research-validated and reliable instrument.

However, these instruments are still being studied. For example, one study found evidence that the FCI is biased against students in Arabic countries (Mealy *et al.*, 2003). Another study on women enrolled in U.S. institutions found that the FCI is biased against women (Traxler *et al.*, 2018). The examples used in the FCI can be culturally offensive or irrelevant, the latter of which can alter scores (McCullough, 2004). This calls into question what the FCI is measuring when it comes to different student populations. These issues are not unique to the FCI. A study on the Colorado Learning Attitudes about Science Survey (CLASS) found that the CLASS was valid for Hispanic students, but interviews with students revealed that one question was not interpreted as intended by CLASS creators (Sawtelle *et al.*, 2009). These biases indicate that these instruments may not be as stalwart as believed. As indicated by Lindell and Ding (2012), validation of assessment instruments in education research should not be a one-time task, but instead it should be an on-going process for ever-changing participants, contexts, and purposes. To mitigate possible cultural biases in the CLASS, researchers have used a transadaption process to translate the CLASS into Mandarin (Zhang and Ding, 2013).

Additionally, as discussed in the previous section, research may not be relevant to the populations of interest or be inclusive of what is being measured. These instruments have limitations in what they measure; thus, a broader area, such as learning physics, may occur but not in the way that the instrument designers defined learning. Thus, carefully describing the findings is important for accurately reporting what is known.

Knowing the background of the research instrument's development, such as for whom was the instrument validated, can help researchers make more equitable decisions. Awareness of the various research being conducted on these instruments that look for bias can also help.

Questions for reflection

1. What data will be collected?
2. What are the research participants being asked to do?
    a. How might questions or tasks be distressing to people with marginalized identities
    b. What is gained by asking questions or doing tasks that could be distressing?
3. What is the validity and reliability evidence for the population of interest?
    a. Has the assessment audience changed to warrant reexamination?
    b. Have other contextual factors changed to warrant reexamination?
4. What is the current research on already developed instruments?
    a. Has the validity and reliability evidence been updated?

5.6 DATA SOURCES AND COLLECTION

Researchers in PER obtain data where participants are aware of data collection such as interviews, observations, and surveys. They also collect data via means that participants are not aware of, such as grades or admission patterns.

5.6.1 Using "big data" and other already existing data sets

As technology evolves and more of our lives are digital, massive amounts of data (i.e., big data) are collected without much additional effort from participants; these are often described as "big data" and employ various machine learning and data mining techniques to uncover findings that may not be obvious. There are also several data repositories [e.g., PhysPort.org; Learning About STEM Student Outcomes (LASSO)] available for those interested in PER at the time of this writing, where individual researchers pool together data sets.

Even before algorithms are run, the data can still have biases toward marginalized people. One example is grades which contain biases (Zeide, 2017). At some point in the grading process, a person decides what is being graded and how to assign points to activities, homework, and exams; some items that could be part of the grade, such as attendance or participation, may not account for students who have no reliable transportation or students who participate in the ways not viewed as active participation by the instructor (e.g., writing answers instead of verbally answering questions). While researchers oppose grades as a metric used in research (e.g., Pearson *et al.*, 2022), big data can still introduce issues. The full context and systemic injustices may not be included in models (Zeide, 2017). There is also a chance that big data PER studies might draw upon wealthy, predominantly white, four-year institutions unless researchers deliberately work with other types of institutions. Thus, using big data in PER *could* be a way to reduce the burden on marginalized research participants for some types of studies but could also reproduce the issues identified by Kanim and Cid (2020).

Lastly, there are multiple concerns regarding big data regarding consent (Johnson, 2014; and Regan and Jess, 2019), privacy (Johnson, 2014; and Regan and Jess, 2019), and how the data are used in education (Johnson, 2014; Zeide, 2017; and Regan and Jess, 2019). Individuals may not be aware that information about them is being shared and used, even though they likely have signed legal documents for their data to be used. Even though big data can mean hundreds or more individuals, studies have shown that a few variables can lead to identifying an individual (Johnson, 2014). As mentioned previously, this could easily happen in physics given the current demographics.

The data can be used in positive manners to better support student learning but also in negative manners such as tracking students and not accounting for potential changes in individuals (Johnson, 2014; Zeide, 2017; and Regan and Jess, 2019). Such data use can reinforce assumptions and stereotypes. Even if researchers are not using data in this manner, some people are aware of these uses and may be uncomfortable because they did not explicitly consent. These concerns are not unique to big data but can be pronounced.

We bring these issues up for researchers to consider, not dissuading researchers from using such data, but to highlight issues that may not be obvious. Big data are attractive given the sheer volume of information available and some novel insights that can be revealed. However, like other types of data and acts of data collection, big data can perpetuate inequitable situations.

5.6.2 Considerations when collecting data

Research settings can support or impede participation. Some settings may be inaccessible for many reasons, including participants not being able to physically enter a space (Chini *et al.*, 2021) or going to locations that might reveal marginalized identities that the participants wish to conceal (Griffiths *et al.*, 2017). For the latter, if interviews take place in a public space and the researchers are well-known, others could see the participant taking part in an interview and guess the participants' marginalized identities. Even for studies that are online, participating in a research study may still not be accessible to all potential participants because internet access is not guaranteed for all people (e.g., Cochran *et al.*, 2021). Thinking through what is required of research participants to participate in a study and identifying potential barriers can help researchers recognize inequities to research participation and, ideally, remove such barriers. This may mean allowing participants to select interview locations (Griffiths *et al.*, 2017) and/or creating multiple ways to participate in research (Aydarova, 2019).

Preparation for data collection is needed to ensure that research participants are treated respectfully. Researchers may inadvertently cause harm by not understanding the traditions or cultural practices of those being researched (Wilson and Neville, 2009). Being aware of various considerations that are culturally appropriate, such as clothing, should occur prior to interviews (Joseph and Ahmed, 2021). For interview data, having a researcher who shares identities with that of the research participant can create comfort for the research participants (Muhammd *et al.*, 2015). When working as a team, researchers should be mindful of the interviewer's workload and ensure they are not doing work disproportionate to their agreed-upon role and compensation, especially if the interviewer has marginalized identities and the interviews will likely be emotionally taxing. While researchers who are familiar with the population may help prevent some missteps, researchers should still engage with prospective research participants to discuss any concerns (Wilson and Neville, 2009). Prospective research participants are ultimately the ones impacted.

The tools and ways that researchers use to collect data can be of equity matter. Filming research participants with marginalized identities can be uncomfortable and make research participants feel as though they should present themselves in a publicly acceptable way that conceals parts of themselves (Gregory, 2020). Researchers can reconceptualize themselves as dialogic listeners and prioritize daily experiences and relationships over more typical research methods that are led by the interviewer (San Pedro and Kinloch, 2017; and San Pedro 2020); these methodical choices are part of indigenizing or humanizing research methods. Research participants may be more comfortable using other forms of expression such as social media posts or selfies (Wargo, 2017).

What is important to note is that the latter two methods are designed to give more agency to research participants regarding what they share. They are also more in alignment with what people might typically do in their lives, blurring the lines of what "counts" as data and data collection. These methods might be a better way of collecting data to answer research questions that are important to the research participants. We are unaware of these or similar methods commonly being used in PER as described in the cited studies, but we anticipate them to be relevant to many studies as they provide more equitable ways of exploring relationships and identity in PER.

Questions for reflection

1. How will data be obtained?
    a. When and where will the data be collected?
    b. What forms of data will be collected?
2. Who is collecting the data?
    a. How does the identity of the data collector relate to the identity of the researched?
    b. Is the workload equitably distributed among the researchers?

3. What barriers exist for research participants to participate?
    a. Are there ways to overcome these barriers?
4. Can pre-existing data answer the research questions of interest?
    a. Are there issues in using pre-existing data?
    b. What measures can be taken to mitigate such issues?

5.7 SAFETY AND CARE FOR RESEARCHERS AND RESEARCH PARTICIPANTS

There are multiple issues to consider around safety and care for researchers and research participants. Like other themes within this chapter, these issues apply to many types of PER work and not just EDI work in PER. Although researchers gain Institutional Research Board (IRB) approval before embarking on studies and are asked to consider risk and harm, some of the risks and harms discussed below may not catch the attention of the IRB. Thus, a study can be IRB approved and still be inequitable to researchers and research participants. These considerations ask researchers to plan beyond IRB compliance in research design to work toward equity.

Given the seriousness of the equity issues, the suggestions may not cover all situations. Seeking out advice or collaboration from researchers who work in disciplines or areas (e.g., social work, gender and sexuality studies) where safety and care considerations are more commonly discussed and integrated into formalized education is valuable for more specific guidance. We anticipate that many of these issues can occur at any stage of the research; therefore, they should be integrated into the research design as appropriate.

Interacting with participants may place researchers in unsafe situations. For fieldwork, recommendations include creating safety protocols such as having members of the research team check-in with one another and think through scenarios such as what they can do and where they might encounter harassment (Sharp and Kremer, 2006). Although researchers in PER may not see their work as similar to fieldwork in other disciplines like geosciences, there are commonalities like travel and secluded settings for interviews. We further encourage research teams to think through what they can do beyond legal compliance to support research participants if the team learns that a researcher has caused harm. Other works (e.g., National Academies, 2018) has advocated similarly, pushing for a culture and climate that is intolerant of harassment.

None of the above planning is to blame anyone for harassment or harm they experience or suggest that researchers can simply plan for safe environments, and all will be well. We acknowledge that these harmful situations are unjust and that incorporating safety measures into the research is not a completely equitable solution. In the discussion section of this chapter, we discuss the importance of working on equity matters in other areas of researchers' lives because they impact research and its design. However, acknowledging the realities of research and being prepared to handle these issues can prevent them from being exacerbated.

Some issues are less deliberate but still can be difficult for researchers to navigate. Researchers who have hidden marginalized identities (e.g., invisible disabilities or identify as LGBTQ+) might feel obligated to disclose those identities to create rapport and trust with the research participants (Secules *et al.*, 2021). Disclosing negative experiences (e.g., harassment, not passing course) may build rapport and establish trust, but sharing such information places the researcher in a vulnerable position. Even if the researchers willingly share marginalized identities or vulnerable experiences with the research team, they may not wish to do so with the research participants. Researchers may particularly be uncomfortable with disclosing their marginalized identities or experiences with research participants they frequently see or if the research participants have power over the researcher (e.g., a postdoc interviewing faculty). Researchers should consider what they are willing to reveal to the research participants and plan accordingly in the research design.

Researchers with marginalized identities may strongly identify with research participants experiencing negative situations, to a point that researchers are negatively impacted (Griffiths *et al.*, 2017; and Kumar and Cavallaro, 2018). Some impacts include emotional struggles, compassion fatigue, where researchers experience physical, mental, emotional, and psychological challenges from helping others, and avoiding research tasks not to be emotionally distressed (Kumar and Cavallaro, 2018). This can occur during the data collection or during analysis. For some studies in PER that are explicitly interested in the experiences of people with marginalized identities, this risk can be anticipated. However, research participants may discuss difficult experiences even when the researcher did not ask. For example, asking a student about their future may reveal that they have a hidden marginalized identity and had difficult experiences that led them to make a particular career choice. The research team should be aware that these situations happen and should include different types of support for researchers in their research design. This may mean identifying institutional resources for support, figuring out workload or a timeline that gives adequate time for researchers not to be overwhelmed with difficult work, and regular check-ins to provide additional support from the research team.

Research participants can have negative experiences while participating in studies. By participating in a study, research participants may be asked to recall difficult experiences (Griffiths *et al.*, 2017). In PER, there are studies focused on people with marginalized identities in physics where participants are asked to share painful experiences. Researchers should be mindful that research participants may be distressed during or after the interview and should provide resources and support after the data collection (Ferreira and Ferreira, 2015; and Griffiths *et al.*, 2017). Although the examples are from interviews, distress could occur if participants are asked about sensitive topics in a survey.

Some suggestions of what researchers can do include offering spaces where research participants can reflect or process emotions (Griffiths *et al.*, 2017), offering lists of resources for participants (Ferreira and Ferreira, 2015; and Griffiths *et al.*, 2017), or having a post-interview discussion after some time has passed (Nardon and Aarma, 2021). Despite these potential difficulties that research participants might experience, interviews can be therapeutic or allow research participants to gain a better understanding of their experiences (Nardon and Aarma, 2021). While all these practices are endorsed by different researchers, not all researchers may have the skillset to avoid exacerbating distress. Seeking out training and advice from others who have more familiarity with these suggestions will be helpful (Nardon and Aarma, 2021).

Questions for reflection

1. In what ways could the data collection process be harmful or traumatic for researchers? For research participants?
2. What safety plans have been created?
   a. What informs these plans?
   b. What mechanisms are in place to ensure the implementation of these plans?
3. What support and care systems are built into the research?
   a. How can researchers be supported?
   b. How can research participants be supported?
4. How can research teams discuss and address non-anticipated challenges when these events occur?

5.8 THEORETICAL FRAMEWORKS, DATA ANALYSIS, AND INTERPRETATION

As analysis approaches are contingent on data, researchers can refine practices of data analysis and interpretation. Carefully constructing plans can help conduct a more equitable research at these later stages by thinking through known and potential challenges.

5.8.1 What theories and definitions can communicate regarding equity

There are many choices during the data analysis stage that can impact equity in research. Researchers select theoretical lenses that guide their research, often selecting popular frameworks such as communities of practice (CoP). Despite widespread use, even well-established theoretical lenses are imperfect. For example, the Tinto model for student persistence has been critiqued by researchers for excluding environmental factors at institutions that can impact student persistence and for assuming that students should culturally assimilate to predominantly white institutions (Museus, 2014). When using theories, researchers should consider what the chosen theories are conveying and whether there are new developments or alternatives that are more equitable. Museus (2014) proposed the Culturally Engaging Campus Environment (CECE) model as an alternative to the Tinto model, incorporating institutional environmental factors in the CECE model.

Variable definitions can also communicate ideas around equity. Rodriguez *et al.* (2012) explored how definitions of equity can change results. Research has shown how aggregating demographic data mask disparities (Museus and Kiang, 2009; Griffiths *et al.*, 2017; and Teranishi and Kim, 2017) and perpetuates stereotypes that harm marginalized communities (Museus and Kiang, 2009; and Teranishi and Kim, 2017). One such example of aggregate demographic data masking disaparitities is the "model minority myth" that perpetuates such stereotypes that Asian Americans have been successful and therefore do not need any support. Data disaggregation has shown that measures of success, such as higher education attainment, vary wildly for different ethnic groups aggregated into the Asian American category (Museus and Kiang, 2009; and Teranishi and Kim, 2017). Sometimes these data are collected as an aggregate category. Other times researchers aggregate categories for a variety of reasons, including achieving a large count to meet the needed statistical power on data. While it may be desirable to obtain a sufficient number of participants for each of the disaggregated groups, considerations must be given to the feasibility of quantitative studies in these cases. This tie back to the issue of research questions and goals mentioned above. Good and sensible research design should not force researchers into adopting quantitative analysis when disaggregated participants are known to be few, while such disaggregation is key to the researchers. In such cases, qualitative design is perhaps a more viable approach.

We further argue that beyond the study results, these decisions can influence how the audience perceives these broad concepts. If researchers define equity meaning equal, practitioners may think equitable practices mean ensuring students have equal amounts of time to communicate in a class discussion, which may not be what students actually need or want. Aggregating categories imply that these identities are similar. Given that aggregating categories can conceal inequities, disaggregated data should be used whenever possible. If it must be done, the theoretical reasons for why aggregate categories are used should be stated and limitations should be noted.

Another challenge researchers may grapple with is analyzing the data and considering intersectionality. Using intersectionality as an analysis lens is not simply considering each demographic identity separately. For quantitative work, novel methods for analysis are needed rather than relying on established and accepted statistical methods that were not designed for intersectional studies (Bowleg, 2008). There are promising methods being developed in PER to analyze quantitative data from an intersectional lens (e.g., Doyle, 2017). For qualitative work, intersectionality as a lens has allowed researchers to examine their own complex identities, which has revealed many previously overlooked issues, such as researchers' tacit assumptions, beliefs, and attitudes, together with their impacts on the researched and on data interpretation can be fully explicated and articulated. This in turn affords richer and more contextualized accounts of marginalized participants, including their struggles and opportunities (Evans-Winters and Esposito, 2019). Although we have separated qualitative and quantitative methods in this paragraph, this divide is somewhat artificial. Bowleg (2008) pointed out that there is considerable overlap in challenges for intersectional studies, regardless of whether the study is qualitative or quantitative. Thus, considering challenges and perhaps solutions broadly in the research may help develop novel solutions.

5.8.2 Reflecting on results

Results can provoke harm although technically being true. They can reify stereotypes if the results do not account for greater contextual factors (Gaddy and Scott, 2020). Building in mechanisms such as engaging with the population studied or the research participants themselves can provide additional insights. As discussed in other sections, there are many ways to obtain such feedback.

Researchers should also be cautious when the sample sizes are small and few people have a marginalized identity. Overgeneralizing the results has been identified as an issue within PER work (Knaub *et al.*, 2019). Clear descriptions of the sample and identifying important limitations are steps that researchers can take to ensure their audience understands the work, but there are other options. Researchers can collect data from more sites to increase the sample size; we do urge caution with this because as discussed earlier, a larger sample does not unnecessarily mean a diverse sample. Depending on the research questions and skillset of the research team, other data may also answer the research questions.

Questions for reflection

1. How may the collected data be analyzed and interpreted in ways that can minimize unintended issues?
2. What messages, explicit or implicit, are being conveyed?
    a. In what possible ways can the intended message be misconstrued?
    b. What measures can be taken to avoid or minimize such mis-construal?
3. What are the potential pitfalls of the analysis techniques being used?
    a. How have other researchers handled this issue?

5.9 SHARING THE RESEARCH FINDINGS BEYOND THE ACADEMY

PER is disseminated through typical academic means such as peer-reviewed journal articles, book chapters, proceedings papers, and talks at conferences or other events. The audience is typically other academics who may be interested in knowing about the findings for purely scholarly reasons. They may also be interested in a particular research technique or in implementing one of the findings.

Sharing research findings with non-academic audiences may not initially seem like a research design matter. However, sharing research findings can and arguably should be part of the research design. Throughout this chapter, a constant thread is the need to be engaged with the population being studied. There is also a thread about the extent to which research participants truly benefit from the research. Sharing findings can be beneficial to research participants, especially if the findings are to be used by the research participants (Aydarova, 2019). For example, physics educators might become aware of equity issues around exclusive, harmful language and work to remove such language from syllabi and assignments.

Like other equity concerns in research, sharing research findings has more nuances. Accessing research can be challenging for people for reasons related to finances (Aydarova, 2019; and Dounas-Frazer *et al.*, 2021), accessibility and disabilities (Dounas-Frazer *et al.*, 2021), and specialized academic jargon (Holmes, 2019). Research can be shared through media aimed at the general audience, such as magazines, podcasts, or blog posts (Aydarova, 2019; and Holmes, 2019). In PER, researchers can create short summaries of PER studies on a site called PERBites (PERBites). Each of these suggestions reaches different audiences and has different constraints. For example, a magazine article likely reaches a larger audience than a post on a researcher's personal blog, but would involve submitting a magazine article and having it be accepted.

Questions for reflection

1. How can research findings be shared to maximize their intended benefits?
    a. Who will see the research findings?
    b. In what settings and formats will findings be shared?
2. How will research participants access the findings?
    a. What barriers were there for the research participants in accessing the findings?
    b. What measures can be taken to remove such barriers?

5.10 CHANGES TO RESEARCH DESIGN

Having a robust research design and considering how the research might perpetuate inequities are key to mitigating harm. Carefully articulating the activities and identifying potential issues can significantly prepare researchers for reducing negative impacts of such issues. However, even with plans that have the most careful thinking, there may be a need to change the research design.

Researchers and research participants are impacted by the broader society. For example, students experienced multiple stressful crises during the Covid-19 pandemic, including economic and racial injustice, health, and access to means of studying at home (Cochran *et al.*, 2021). Individuals can be directly or indirectly impacted by challenging societal events (Ferreira and Ferrira, 2015; and Cochran *et al.*, 2021). Experiencing these injustices and stressful events put additional strain on research participants. Participating in research projects during a crisis may not be desired, especially if the research is on emotionally difficult topics.

There are some strategies to change the research design if the research may be or is harmful. Thinking through potential unintended, harmful consequences with research participants before embarking on a project and thinking through what could be changed to mitigate harm is one possibility (Oliver and Tinkler, 2020). Determining what could be done at a more relaxed pace is another option (Little, 2020). Some of these decisions can be made by examining the research goals (Little, 2020; and Oliver and Tinkler, 2020). Ending a project entirely is also an option if researchers believe that their work is causing more harm than good (Gregory, 2020).

These decisions can impact more than just whether a study is published. Little (2020) noted the potential financial and other life impacts on researchers. Doctoral students, for example, need to conduct research to graduate with their doctorates. As these situations are idiosyncratic, being able to have honest, compassionate conversations and being open to creative decisions can help researchers make better decisions when changes are needed.

Questions for reflection

1. What mechanisms can help identify any unintended consequences or unanticipated difficulties to research participants and the researchers?
2. In case of unintended consequences, how can researchers address or redress harm?
   a. What can be done to prevent such consequences from happening again?
   b. What will be done if harm cannot be addressed or redressed easily?
3. If there are any unanticipated difficulties, what changes are possible?
   a. How might these changes impact the research participants? Researchers? The findings of the study?
   b. What can be learned from the experience that can inform future research practices?

5.11 DISCUSSION AND CONCLUSIONS

Equity in research designs for PER is an important area, given the power that research has. Equity concerns cover a wide range of research areas in PER. This power can either create a more equitable physics spaces or perpetuate harms leading to inequities. Research design is an important step to make our research more equitable and thus make physics spaces more equitable by designing practices and policies that work for marginalized people.

We recognize that the chapter does not provide researchers who wish to be more equitable in their PER work with exact steps to follow. Much of the consideration of equity in research design is contextual. Being mindful of the importance of context, we do not offer prescriptive solutions that are likely unapplicable to most contexts. Understanding good or “best” practices is bound to the information available at the time as well as the bandwidth of researchers. This is not to say that researchers should not aim to be more equitable, just acknowledging real constraints. Instead, we aim to highlight that equity considerations do require effort from researchers.

Our goals are to highlight some equity issues that arise or could arise in PER studies and encourage researchers to consider how they can include equity-related considerations in their research designs. Planning, rather than leaving decisions to chance, can support researchers in being more equitable. In addition to the reflection questions in each section, we offer the following recommendations:

- **Form positive, respectful relationships with diverse populations.** The literature in this chapter had a consistent thread around relationship building. If researchers in PER are seeking to rectify

PER's history of studies taking place at predominantly white, privileged institutions and conduct research on more diverse physics settings, they need to consider how to do this with equity in mind. Otherwise, such studies could perpetuate or even exacerbate inequities (e.g., propagating stereotypes, treating research participants poorly). Building positive, respectful relationships that *may* grant access to populations of interest cannot be done transactionally, though. We emphasize "may" because having positive relationships does not make any researcher entitled to having access to a population for pure research reasons.

- **Continue learning and improving**. Equitable practices are constantly changing with time as more voices are part of the conversation. Even if researchers are using the latest "best" practices, what are "best" practices today might not be "good" tomorrow. Even if one does not consider themselves a researcher on EDI matters, staying up to date, for example, through short summaries or attending relevant talks can help researchers stay updated.

These might be seen as generic advice, not particular to PER or research design. However, there is a human aspect to everything we do as researchers in PER.

Lastly, as a research community, we need to consider our actions in a broader society. Many of the proposed practices to be incorporated into the research design may not be available for all researchers; revealing marginalized identities is one example of a practice that may not be available for all researchers due to stigma or other consequences. For prospective research participants with marginalized identities, participation may not be worth the risks. Researchers in PER work toward improving the conditions and experiences for those in physics learning spaces. We know that people bring their life experiences to these learning spaces, so a part of improving the conditions and experiences of those in physics learning spaces must include holistically improving people with marginalized identities' conditions and experiences in a broader society.

Although we are addressing the broader societal context with research design in mind, we see more equitable research practices in PER as a fringe benefit to dismantling broader societal oppression. It is advocated that researchers interested in equity become activists or allies in fighting oppression (e.g., Holmes, 2019; Evans-Winters and Esposito, 2019; and Collins, 2021). This can involve working with activist groups in areas of interest (e.g., Holmes, 2019) or disrupting oppressive actions whenever one can (e.g., Collins, 2021).

By centering equity in our lives and working to end structures that create barriers for people with marginalized identities, we can create a more equitable world and be more equitable researchers. This, in turn, impacts physics spaces and conducting equitable research in physics education.

REFERENCES

AIP, (2018), see https://www.aip.org/statistics/reports/beyond-representation-data-improve-situation-women-and-minorities-physics-and.
Anteneodo, C. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010136 (2020).

Aydarova, E., *Research Methods for Social Justice and Equity in Education* (Palgrave Macmillan, Cham, 2019), pp. 33–43.

Boveda, M. and Weinberg, A. E., Phys. Teach. **58**(7), 480–483 (2020).

Bowleg, L., Sex Roles **59**(5), 312–325 (2008).

Chapman, C. and Ainscow, M., Brit. Educ. Res. J. **45**(5), 899–917 (2019).

Chini, J. *et al.*, Doing physics education research inclusively: Designing for variation in participants' needs, abilities, and interests (2021), see https://www.compadre.org/per/perc/2021/Detail.cfm?ID=8392 (last accessed March 8, 2022).

Cochran, G., Guest Post: The Problem with Diversity, Inclusion, and Equity (2018). [online] The Scholarly Kitchen, see https://scholarlykitchen.sspnet.org/2018/06/22/problem-diversity-inclusion-equity/_ (last accessed March 1, 2022).

Cochran, G. L. *et al.*, *2021 ASEE Virtual Annual Conference Content Access* (ASEE, 2021).

Collins, J., Paper presented at *Physics Education Research Conference 2021, Virtual Conference* (PERC, 2021, August 4–5), see https://www.compadre.org/Repository/document/ServeFile.cfm?ID=15804&DocID=5531.

CSSP, *Key Equity Terms and Concepts: A Glossary for Shared Understanding* (Center for the Study of Social Policy, Washington, DC, 2019), see: https://cssp.org/resource/key-equity-terms-concepts/.

D'Ambrosio, B. *et al.*, J. Res. Math. Educ. **44**(1), 11–22 (2013).

Dounas-Frazer, D. R. *et al.*, Committee on Laboratories Accessible Physics Labs Task Force Report (American Association of Physics Teachers, College Park, MD, 2021).

Doyle, J., FIU Electronic Theses and Dissertations (2017), p. 3353.

Evans-Winters, V. E. and Esposito, J., *Qualitative Inquiry at a Crossroads* (Routledge, 2019), pp. 52–64.

Ferreira, R. J. *et al.*, J. Soc. Work Values Ethics **12**(1), 29–40 (2015).

Gaddy, M. and Scott, K., *Principles for Advancing Equitable Data Practice* (Urban Institute, Washington, DC, 2020).

Gregory, K., Int. J. Qualitative Methods **19**, 1609406920963761 (2020).

Griffith, C. *et al.*, J. LGBT Issues Counseling **11**(4), 212–229 (2017).

Holmes, IV, O., Equality, Diversity Inclusion: Int. J. **38**, 668–675 (2019).

Johnson, J. A., Int. Rev. Inf. Ethics **21**, 3–10 (2014).

Joseph, F. I. *et al.*, Int. J. Qualitative Methods **20**, 16094069211058616 (2021).

Kanim, S. and Cid, X. C., Phys. Rev. Phys. Educ. Res. **16**(2), 020106 (2020).

Knaub, A. V. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020001 (2019a).

Knaub, A. V. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020102 (2019b).

Kumar, S. and Cavallaro, L., Qualitative Health Res. **28**(4), 648–658 (2018).

Lindell, R. and Ding, L., *2012 Physics Education Research Conference Proceedings*, edited by P. Engelhardt *et al.* (American Institute of Physics, Melville, NY, 2012), Vol. 1513, pp. 27–29.

Little, A., Managing Grant Team Stress and Looking Out for One Another During COVID — Angela Little LLC. [online] Angela Little LLC (2020), see http://angielittle.com/blog/2020/7/30/managing-grant-team-stress-and-looking-out-for-one-another-during-covid (last accessed March 6, 2022).

McCullough, L., J. Int. Women's Stud. **5**(4), 20–30 (2004).

McDermid, F. *et al.*, Nurse Res. **21**(5), 28–33 (2014).

Mealy, C. *et al.*, *Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition* (ASEE, 2003), pp. 8.350.1–8.350.12.

Mertens, D. M., Qualitative Inquiry **16**(6), 469–474 (2010).

Mertens, D. M. and Ginsberg, P. E., Qualitative Social Work **7**(4), 484–503 (2008).

Muhammad, M. *et al.*, Crit. Sociol. **41**(7–8), 1045–1063 (2015).

Museus, S. D., *Higher Education: Handbook of Theory and Research* (Springer, Dordrecht, 2014), pp. 189–227.

Museus, S. D. and Kiang, P. N., New Directions Inst. Res. **2009**(142), 5–15.

Nardon, L. *et al.*, Int. J. Qualitative Methods **20**, 16094069211065233 (2021).

National Academies of Sciences, Engineering, and Medicine, *Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine* (The National Academies Press, Washington, DC, 2018).

National Research Council, *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Education* (National Academies Press, Washington, DC, 2012).

Ndimande, B. S., Qualitative Inquiry **18**(3), 215–226 (2012).

Oliver, K. *et al.*, Evaluation **26**(1), 61–75 (2020).

Parson, L., *Research Methods for Social Justice and Equity in Education* (Palgrave Macmillan, Cham, 2019), pp. 15–32.

Pearson, M. I. *et al.*, CBE—Life Sci. Educ. **21**(1), 1–10 (2022).

PERBites. https://perbites.org/ (last accessed March 6, 2022).

Pezalla, A. E. *et al.*, Qualitative Res. **12**(2), 165–185 (2012).

Plowman, D. A. *et al.*, Leadership Quarterly **18**(4), 341–356 (2007).

Quinn, C. R., Health Justice **3**(1), 1–7 (2015).

Regan, P. M. and Jesse, J., Ethics Inf. Technol. **21**(3), 167–179 (2019).

Rodriguez, I. *et al.*, Phys. Rev. Spec. Top.-Phys. Educ. Res. **8**(2), 020103 (2012).

Sandoval, C. D. M. *et al.*, Alternative: Int. J. Indig. Peoples **12**(1), 18–31 (2016).

San Pedro, T., *Protecting the Promise: Indigenous Education Between Mothers and Their Children* (Teachers College Press, 2021).

San Pedro, T. and Kinloch, V., Am. Educ. Res. J. **54**, 373–394 (2017).
Sawtelle, V. *et al.*, Phys. Rev. Phys. Educ. Res. **5**(2), 023101 (2009).
Secules, S. *et al.*, J. Eng. Educ. **110**(1), 19–43 (2021).
Sharp, G. and Kremer, E., Sociol. Methodol. **36**(1), 317–327 (2006).
Sohr, E. R. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020157 (2020).
Stahl, G. K. *et al.*, J. Int. Business Stud. **41**(4), 690–709 (2010).
Strunk, K. K. and Hoover, P. D., *Research Methods for Social Justice and Equity in Education* (Palgrave Macmillan, Cham, 2019), pp. 191–201.
Teranishi, R. T. and Kim, V., Edu. Forum 81, 204–216 (2017).
The Annie E. Casey Foundation, Equity vs. Equality and Other Racial Justice Definitions (2022) [online], see https://www.aecf.org/blog/racial-justice-definitions (last accessed March 6, 2022).
Traxler, A. L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020114 (2016).
Traxler, A. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(1), 010103 (2018).
Uhl-Bien, M. *et al.*, Leadership Quarterly **18**(4), 298–318 (2007).
Wargo, J. M., New Media Soc. **19**(4), 560–578 (2017).
Wilson, D. and Neville, S., Contemp. Nurse **33**(1), 69–79 (2009).
Zeide, E., Big Data **5**(2), 164–172 (2017).
Zhang, B. *et al.*, Int. J. Qualitative Methods **20**, 16094069211064672 (2021).
Zhang, P. and Ding, L., Phys. Rev. Phys. Educ. Res. Phys. **9**(1), 010110 (2013).

SECTION
II
THE HISTORY AND PHILOSOPHY OF PHYSICS IN PHYSICS EDUCATION

Section Editors

Donald J. Metz, Peter Heering, and Cibelle Celestino Silva

CHAPTER
6 PHYSICS AS A HUMAN ENDEAVOR

Richard Staley

Staley, R., “Physics as a human endeavor,” in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 6-1–6-26.

6.1 INTRODUCTION

What difference does it make to think of physics as a *human* endeavor, rather than as we normally might, as a complex of conceptual insights and principles, predictive theories, experimental machines, and sophisticated technologies? The relevant contrast case is probably not with physics as a non-human endeavor. Considering all the material, non-human elements of some representative physical systems like space-based telescopes, missile defence systems, gravity wave observatories, particle accelerators or even teaching experiments are likely instead to remind us of the extraordinary social and financial resources that each requires to deliver their fruits. Rather, we might wonder what the distance is between physics pursued with human and beyond-human aims. In the early twentieth century the great German theoretical physicist Max Planck offered an influential ideal image of the progress of physics, describing his science as offering a continual deanthropomorphization, sacrificing the incommunicable experience of color for measurable wavelength, and arguing that rather than seeking the adjustment of thoughts to perception, physics reached above all for the complete liberation of the physical world picture from the individuality of the creative mind. The physical world picture was independent of the good will of individual scientists, independent of nationalities and centuries, and even independent of the human species itself. Based on absolute constants, general principles, and the invariants of the theory of relativity, it expressed a unity concerning all places and times, all scientists, nations, and cultures (Planck, 1909; and Heilbron, 2000, pp. 44–60). Yet in the aftermath of World War I, Planck’s community was riven by a set of intersecting conflicts between theorists and experimentalists, Berlin and provincial physicists, and the articulation—and vicious use—of concepts of Jewish and German physics; while recent studies of Black physicists’ identity in the United States have found that the argument that physics has a “culture of no culture” and the belief that objective physics is beyond bias continues to ring false, imposing barriers to groups who have been excluded

https://doi.org/10.1063/9780735425514_006

or marginalized and are too well aware of the biases of powerful elites in physics (Hyater-Adams *et al.*, 2018, 2019). Emphasizing the humanity of physics has therefore become a way of insisting that physics is pursued by people with a rich diversity of backgrounds differently enmeshed in social and cultural contexts, and also engaged with a rich range of ideological, political, economic and religious commitments. It might be traced back to the egalitarian spirit of the 1948 United Nations "Universal Declaration of Human Rights," which in its 26th and 27th articles asserts the rights of everyone to education, and of everyone to share in scientific advancement and its benefits. Indeed, it is now the absence of significant groups from physics classrooms as much as the wish to teach better those who are already there that drive recent work to deliver a richly human physics.

This chapter opens by showing that those who pioneered physics education research in the 1970s, like Lillian McDermott working at the University of Washington, were often motivated by difficulties of access and responded to cultural and demographic limitations, even if they focused almost exclusively on improving student understanding of classroom content. For that reason, treatments of the nature of science and science as a process provided the primary pedagogical framework for early work to broaden perspectives and show both the creativity of work in physics, and how physics can better serve academically disadvantaged students (most from minority ethnic backgrounds). Responding to a felt need to broaden the base for participation in Science, Technology and Medicine, in 2013 the U.S. National Research Council provided Next Generation Science Standards that explicitly incorporated curriculum goals conveying "science as a human endeavor." While noting the significance of gender and different backgrounds, and considering the mutual interrelations between technology, science and society, I show that these standards might nevertheless reinforce common assumptions about the boundaries between what counts as scientific and cultural, and about the spheres in which ethical questions might be raised. My second and third sections address physics education research on gender, race, and intersectional approaches. Drawing attention to its recent relative isolation from the physics discipline and general educational research, I show that while physics communities had highlighted the poor participation of women from the 1970s and educational researchers in the United Kingdom had developed valuably comprehensive approaches as early as the 1980s, it was only from the late 1990s that U.S. members of the emerging subdiscipline of physics education research devoted significant attention to the barriers to participation and interest and achievement gaps experienced by women and racial minorities. Recent work has addressed bias in such significant pedagogical tools as the force concept inventory and offered richly comprehensive treatments of the lived experience of minority scientists. Focusing on identity performance rather than preconceived categories might also yield valuable insights for teaching, learning and policy initiatives. The fourth section develops historical and anthropological resources to understand why participation beyond white male elites has proven such a longstanding limitation, and to assist the development of methodological perspectives on the intersections of the intellectual, cultural and material dimensions that must be addressed to achieve substantial changes in physics education and research. Finally, the chapter concludes by considering the broadest framework in which scientists have raised questions of human responsibility. Humanity has marked the earth so thoroughly that geologists now recognize a new epoch in the Anthropocene. Similarly, global warming has become such an urgent issue that it requires the collective work of humanity to prevent those changes from damaging the earth irrevocably. Understanding and addressing climate change relies on atmospheric physics, ocean circulations, the carbon cycle and cloud dynamics and raises more widely distributed social, technical and economic challenges than high-energy physics, quantum devices or materials science. Asking how physics education might meet future needs in this regard provides a further means of establishing limits to our current understanding of physics as a human endeavor, and suggests important new goals.

6.2 EARLY PHYSICS EDUCATION RESEARCH: DIVERSITY, INCLUSION, AND THE LIMITS OF THE PHYSICS CLASSROOM

Recent commentators have stressed again and again a remarkable feature of the demographic makeup of the physics discipline that has in fact remained similar throughout its history in western countries. This is the dominance of males and cultural elites. Jim Megaw's 1989–90 survey of the world's physics departments showed that industrialized western nations with strong physics establishments (and women's rights movements) had amongst the lowest proportions of women faculty, with departments in Japan, Canada, West Germany, Switzerland, Norway, Korea, the U.S.A. and U.K. all reporting fewer than 5% women faculty, while some Latin and communist countries reported between 16% and 34% (France, Spain, Poland, Brazil, Türkiye, Italy, Thailand, the USSR, the Philippines and Portugal), and 46% of Hungarian physics faculty were women (Megaw, 1992; Baringa 1994; and Götschel, 2011). Considering the United States can provide a more detailed breakdown, that shows some recent changes. Summarizing recent statistical studies from the American Institute of Physics (which also track doctoral demographics from the 1920s), Miguel Rodriguez and colleagues note that at present women make up about 20% of the field at all levels, from bachelor degrees to faculty positions, and that while people of color earn 23% of bachelor degrees and 16% of doctorates, only 6.5% of faculty positions are held by non-white and non-Asian scholars (Ivie *et al.*, 2014; Porter and Ivie, 2019; Mulvey and Nicholson, 2020; Mulvey *et al.*, 2021; and Rodriguez *et al.*, 2022, p. 1). Physics education researchers currently see this as a major difficulty to be overcome, and while they trace the emergence of their discipline to a transformation of physics education that began to incorporate systematic research into its effectiveness from the 1970s, they commonly date the onset of serious concern with women and minority participants only to the late 1990s. Laura McCullough's recent overview of research on gender and under-represented minorities in physics provides an example, identifying the first and second articles on women in physics published in *American Journal of Physics* (the principal journal of the Association of American Physics Teachers) as appearing sixteen years apart in 1976 and 1992. She reports further that the first article on under-represented minorities they published appeared only in 1999, a study of introductory courses taught at Rutgers University (Moché, 1976; Heller and Hollabaugh, 1992; Etkina *et al.*, 1999; and McCullough, 2018). As well as limiting her inquiry to the most longstanding journal that specialized in physics education, McCullough's view reflects a tight focus on methodology, associating physics education research with the production of pedagogical materials and tests of their effectiveness, in particular. This was certainly the direction that pioneering work in the University of Washington had taken when in the early 1970s Lillian McDermott joined Arnold Arons in founding the first physics education group devoted to research on physics learning and teaching (Beichner, 2009; and Meltzer and Otero, 2015). Yet briefly describing the circumstances in which McDermott began this work will provide insight into the mix of social and conceptual concerns characteristic of research in this early phase in the emergence of the discipline—and suggest that in fact in significant respects this work was a response to the cultural and demographic limitations of physics in the United States, even if this was addressed by focusing almost exclusively on improving student understanding of classroom content.

McDermott followed her husband (also a physicist) to Seattle after a postdoc at the University of Illinois and a year teaching physics at the City University of New York. She had herself attended an all-girl high school in New York and the all-women Vassar College before gaining a Ph.D. in nuclear physics at Columbia University in 1959. McDermott later recalled that the Nobel Laureate I.I. Rabi had told her entering class in 1952 that one half of them would be gone by the end of the year. In an oral history interview, she commented that Columbia "was not a very kind place" (McDermott, 2020, 2021, pp. 1–3). Initially assisting as a volunteer in courses for prospective elementary and high school teachers as well as university courses in physics with Arons at the University of Washington, McDermott won a tenure track position in physics education there in 1973 (when nepotism rules were relaxed); and helped initiate a new focus on testing learning rather than on curriculum design. This complemented Arons's concern with teaching physics through inquiry, and her group also linked physics education research with other disciplinary approaches such as developmental psychology. Working with David Trowbridge in the early 1980s McDermott offered treatments of velocity and acceleration influenced by Jean Piaget's work on developmental stages; Beichner writes that these are "recognized as the first 'modern' PER papers" (Trowbridge and McDermott, 1980, 1981; and Beichner, 2009). McDermott's group were also pioneers in addressing the difficulties met by academically disadvantaged students attending introductory courses in the sciences in the Educational Opportunity Program at the University of Washington, noting that most were from ethnic minorities. They observed that academically disadvantaged students suffered from a lack of preparation but more specifically from a battery of factors: a lack of experience on which to base scientific abstractions, weakness in mathematical and verbal skills, lack of confidence in their own reasoning abilities, and low personal standards for academic achievement. Singly, any one of these would be devastating. Appearing together, McDermott's group argued that they also needed to be addressed in concert. In response, they developed courses designed for minority students and offering a comprehensive approach that attended to the course curriculum, the way it is taught, and supporting factors—including assistance in continuing the course when it becomes difficult (McDermott *et al.*, 1980, 1983). We can see that the endeavor was in fact very similar to that on which Rutgers physicists reported two decades later, but it escaped McCullough's notice because it was published in the *Journal for College Science Teaching*, befitting the authors' wider focus on introductory courses in biology as well as physics. While these papers clearly establish the breadth and social awareness of McDermott's group, it is also evident that McDermott focused particularly on understanding the relations (and testing the gap) between teaching and learning, developing iterative methods to disclose conceptual difficulties. Throughout her career, McDermott's research publications have left gender issues largely untouched.

This was characteristic of physics education research as a whole in this period: the instruments that enabled the rise of this subfield within the physics discipline were primarily tests through which physicists could probe the conceptual understanding of their students, such as, in particular, the Force Concept Inventory developed by David Hestenes, Malcolm Wells, and Gregg Swackhamer in 1992. This offered forced choices between Newtonian concepts of force and common sense beliefs—usually labeled misconceptions, but which they argued needed to be taken as seriously as scientific concepts. The most remarkable feature of the tests they had run for 1500 high school and university students was its ability to identify and diagnose a variety of misconceptions, and thereby to evaluate the effectiveness of instruction on the concept of force. The authors did take up demographic issues. While commenting on ethnic backgrounds and immigrant populations, they found no correlation with the rough measure of socioeconomic status they had used based on location, but a strong variation between schools, noting also that the subset of students taking physics "is usually not typical of the student population at the school" (Hestenes *et al.*, 1992). Most early work in physics education research then was firmly centred within physics classrooms. However, over time, the exceptional nature of their student populations—and of the demographic composition of the academic profession of physics at all levels—became an ever more pressing issue for physics education researchers, in particular, because these issues were incorporated in broader research and reform efforts that focused on the nature of science. These were initially fielded within general science education research but have increasingly been incorporated in physics education research, and helped frame questions of diversity and inclusion as part of what it means to understand physics as a human endeavor.

In 2013, the National Research Council of the United States set out what they described as Next Generation Science Standards (NGSS), written by the States for the States. Although directed at secondary school education and science in general, their considerations can help orient our concern with physics education research because they offered a rarely explicit treatment of "science as a human endeavor," and the framing concepts that they articulated illustrate tensions that are also characteristic of the way human and social elements are considered in physics classrooms across all levels of education. Their explicit focus on such framing concerns reflected a view that the demographic profile of science education was changing rapidly, posing new challenges in recognizing diversity in the classroom, and that science education should address the nature of science as well as conveying the practices and content of science, but also be widened to incorporate engineering on the same footing as science—something they thought could help broaden access to students traditionally underserved. Although certainly timely in 2013, similar points were made from the last quarter of the nineteenth century when higher education in Europe and the United States began incorporating practical engineering training and opening up to women (with the latter change occurring still more gradually and with considerable protest).

In addition to describing Science and Engineering Practices such as using, developing and synthesizing models, and Disciplinary Core Ideas such as the structure and properties of matter, and organizing these both by core ideas and by topic at different levels of secondary education, the NGSS also set out Cross-Cutting Concepts such as patterns evident across different scales. It is within these that they explicitly noted respects in which the sciences should be understood as "a human endeavor" (National Research Council, 2013a, 2013b). As well as their overarching recognition that science is a result of endeavor, imagination and creativity, bullet points highlight the mutual interrelations between scientific content and the contexts in which it is developed, in formulations appropriate for different educational stages. For example, for early years from Kindergarten through the two stages of Middle School, the Standards note explicitly that "men and women of diverse backgrounds/from all cultures and backgrounds / from different social, cultural, and ethnic backgrounds" work as scientists and engineers. Gender then drops away from NGSS formulations for high school standards, where they note that "individuals and teams from different cultures and backgrounds" have contributed to science, and also make the important but more subtle point that the nature of scientific findings "is influenced by scientists' backgrounds, theoretical commitments and fields of endeavor." Similarly, the NGSS asks educators to convey the mutual interrelationships between technology, society and science: while technological advances have influenced the progress of science, science has also influenced advances in technology; likewise, while science and engineering are influenced by society, society is influenced by science and engineering (National Research Council, 2013b, Appendix H, p. 100). Recognizing these interrelationships surely conveys a sense of responsibility to the social environment, but it is revealing that this is also subtly varied in relation to different aspects of the physical sciences. For example, considering the core disciplinary ideas of both Energy, and Waves and Their Applications in Technologies for Information Transfer, the Standards point towards the significance of major technological systems and describe scientists (in rather abstract terms) as attempting to increase benefits and decrease costs and risks. However, when discussing the performance requirements of the earth and space sciences, the Standards go two further revealing steps. First, they explicitly note *deep and unanticipated* impacts when considering Earth Systems, and secondly they engage in the question of *ethics* in relation to both the Earth and Human Activities, and later Human Sustainability. In these sections, however, at the very point that significant complexity is recognized in the way science addresses questions about the natural and material world, the Standards also introduce a form of boundary work—noting that science and technology may raise ethical issues for which science, by itself, does not provide answers and solutions. Thus, while scientific knowledge is thought to indicate the horizon of what *can* happen in natural systems, understanding what *should* happen 'involves ethics, values, and human decisions about the use of knowledge, "with many decisions made not using science alone relying on social and cultural contexts for their resolution" (National Research Council, 2013a, pp. 127, 288). These are important points and I want to draw out two different implications. First, it is historically revealing that the physical sciences—at least in regard to matter and energy considerations—are not now thought to engage the complexities of human endeavor in the same way that earth systems and the climate sciences do. Yet as well as noting that similarly sharp ethical and geopolitical questions were in fact characteristic both in the period in which the physics discipline expanded after World War II given the promise and threat of nuclear physics, and in rather different ways in the earlier development of thermodynamics, electromagnetism and energy physics, we will surely want to resist narrow definitions of the physical sciences that exclude their engagement in geophysics and climate. Secondly, we should observe the significant but ambiguous senses in which social or cultural contexts are often construed in some contrast to science. The ambiguous, shifting grounds of what counts as scientific, and what are thought to raise ethical questions—and whether science or scientists play a role in addressing them—raise some of the most challenging and significant features of the questions surrounding how to treat science as a human endeavor; ultimately, it is likely to be most effective to integrate them into core content teaching as well as treating them in their own right.

In emphasizing engineering, the NGSS consciously sought to (re)define the epistemology of science and counter the limited or distorted view that previous standards (and, in particular, those represented by *Science for All Americans* in 1989) had provided by defining science in terms of Western science without representing its historical relations with other cultures. They were responding to a historic increase in student diversity in the classroom (which has also been characteristic of many nations in the twenty first century despite periods of increasingly nationalist focus in both the United States and the United Kingdom), noting that persistent gaps in science achievement remained. Noting that the key points of advantage concern social prestige and institutional advantage rather than numerical majority, the NGSS identified seven “non-dominant” student groups: economically disadvantaged students, racial or ethnic minorities, students with disabilities, English language learners, girls, students in alternative education programs, and gifted and talented students. Physics education research is rarely concerned with targeting science teaching to all students, and physics teachers (especially in higher levels) often seem to be more concerned with filtering out gifted students for disciplinary development rather than addressing all abilities; of these diverse axes of concern, questions of gender have dominated. We will turn to consider this shortly, but first it will be helpful to note one of the ways in which discussions of the nature of science have been important within physics education research in particular.

Although a principal concern in science pedagogy has been to convey and assess specific conceptual and technical claims and methods of problem solving, from the 1950s and 1960s at least, educators have also argued that imparting a more general understanding of the character of scientific work, research goals and practices are central to teaching, and delivered tests of this kind of knowledge (Cooley and Klopfer, 1961; and Lederman *et al.*, 1998). In 1990, the American Association for the Advancement of Science also made this an explicit feature of their recommendations on what particular knowledge of the way science works is required for scientific literacy. The opening chapter of *Science for All Americans* was entitled “Nature of Science,” and was divided into sections on the scientific world view, methods of scientific inquiry and science as an enterprise (Rutherford and Ahlgren, 1991). Since then, a consensus view of the nature of science has emerged that features “human” dimensions as one strand amongst several. In the Australian curriculum, for example, strands imparting science inquiry skills and scientific understanding are regarded as intertwined with the strand on science as a human endeavor (Australian Curriculum, 2022). This is commonly taken to reflect the dynamic nature of science, as a field subject to change and uncertainty, with the continual reassessment of predictions and explanations that are often assisted by technology. A second principal feature is the recognition that the varied uses of science involve complex interactions with a wide range of social, economic, ethical and cultural factors. In each unit, such as on Linear motion and waves or on Revolutions in modern physics, a set of content descriptions framed in terms of science as a human endeavor is designed to explore the complex ways science interacts with society across diverse subjects and in different contexts.

Engagement with the nature of science in physics education research has drawn on the history and philosophy of science and centered on recognizing the diverse ways in which different elements of the nature of science must be understood. A primary example is the work of Norm Lederman and colleagues from the 1990s to the present, aiming to develop instruments capable of reflecting both the consensus view on many epistemological features of the science—such as the theory dependence of observation—while exposing students to a variety of ways in which this can be understood (Lederman *et al.*, 2002, 2014, and 1998). On their account, "scientific knowledge is tentative; empirical; theory-laden; partly the product of human inference, imagination, and creativity; and socially and culturally embedded," and rather than simply focusing on the process of science such as observation, the collection and interpretation of data and the derivation of conclusions, it concerns the values and assumptions underlying these processes (Lederman *et al.*, 2002, p. 499). It is in referring to its social and cultural foundations and creativity that the nature of science touches most clearly on human elements, and to some extent both are seen to run counter to common perceptions of science. But much of Lederman and colleagues' work stemmed rather from a suspicion that in the variety of forced choice tests that had been developed in earlier research, the inability of the instruments to detect variations between the testers' and respondents' understandings of the questions resulted in the very real prospect that irrespective of the choices made, the developers' views were being imposed on respondents (which led to them being interpreted as holding distinct philosophical stances, such as hypothetico-deductivist). Drawing on a Views on Science–Technology–Society (VOSTS) questionnaire developed by Aikenhead *et al.* (1989) and Lederman and O'Malley (1990) instead used an open ended questionnaire which they followed with interviews. The interviews enabled researchers to test inferences they had drawn from responses (and in this early iteration disclosed respondents' difficulties in interpreting three of 7 questions) as well as to assess both respondents positions and why they held particular views. It is revealing that in a doctoral dissertation in which Randy Bell compared novice and expert views (using a later version of the test) in order to test its construct validity, Bell found that novices typically underestimated the creativity of scientific work, thought there to be a single scientific method, and rarely made reference to social or cultural influences on the development of scientific knowledge—a finding that underlines the significance of incorporating nature of science within the requirements of scientific literacy for non-scientists (Bell, 1999; and Lederman *et al.*, 2002, p. 505). By 2002, three forms of the Views of Nature of Science Questionnaire had been administered to about 2000 high school students, college undergraduates and graduates, and preservice and inservice elementary and secondary science teachers across four continents, in company with about 500 individual interviews. The results of these studies and follow-up interviews support a high confidence level in the validity of the VNOS for assessing the NOS understandings of a wide variety of respondents.

How should perspectives on the nature of science be developed in physics classrooms in particular? The Perimeter Institute has developed a teacher guide on the process of science in physics (Fish *et al.*, 2013), and in later sections I will suggest some valuable historical resources that can aid in developing more social and cultural perspectives on the physics discipline. Today, many theoretical and empirical studies use physics to illustrate how we can understand science as a human endeavor. Examples include Hadzigeorgiou's treatment of narrative storytelling in electricity, Kapon and colleagues' discussion of disciplinary authority and personal relevance, and a study of teaching general relativity in South Korea (Hadzigeorgiou, 2006; Kapon *et al.*, 2018; and Park *et al.*, 2019). Galili (2019) observes the central contrast that while scientists learn the nature of science from the inside throughout their professional lives, introductory science education attempts to convey knowledge of the same subject from the outside. He offers a valuable review of the nature of science understandings for education in the physical sciences, arguing that teachers should display a representative variation of views. The nature of science also provides an important platform for considering the identity of physicists. In a recent study of underrepresented groups, Moses Rifkin (2016, p. 73) persuasively argues, "if […] we view physics as a process by which knowledge is uncovered, then we must talk about the 'who' of physics as we are talking about the 'what.'"

6.3 COMBATTING BIAS: GENDER

Two special issues published 29 years apart give a good sense of just how difficult it has been to make progress in addressing gender imbalances in the physics discipline and its teaching. The first was presented in 1987 but has been largely forgotten, although it offered a very similar basic orientation to the Focused Collection that *Physics Education Research* published on Gender in Physics in August 2016. The latter is already doing a great deal to shape further research and has helped justify the contention that although gender has long haunted the physics discipline without attracting the concerted action it deserves, physics education researchers are now developing creative perspectives on it in many different ways. I noted the long distance between the first and second articles on women in physics published in the *American Journal of Physics*. In their incisive 2016 interpretive overview of research methods on gender, Adrienne Traxler and colleagues highlight the second of these and describe it as "the first paper addressed to physicists that included research results regarding gender differences in the physics classroom" (Traxler *et al.*, 2016, p. 2). The paper had found that problem solving groups with two women and one man outperformed those with two men and one woman, even when the single woman involved was the highest-ability member because the men might simply ignore her arguments, correct though they were (Heller and Hollabaugh, 1992, p. 641). Yet leaving the point about just how recently this was published without noting the broader context in America and elsewhere would risk underestimating the scale of the problem—for the fact that so little was published in the *American Journal of Physics* is a clear indication of the extent to which American physics education researchers were insulated from critical aspects of the physics discipline, and educational research elsewhere.

Sustained concerns with women in physics and efforts to redress gender imbalances had begun to emerge a generation earlier in the 1970s, and by the 1980s had led to serious research on education, at least in Britain. The American Physics Society founded a Committee on the Status of Women in Physics in 1972, aiming to "address the encouragement and career development of women physicists" and in Britain the Joint Physics Education Committee of The Royal Society and The Institute of Physics published a Report on *Girls and Physics* in 1982 (Joint Physics Education, 1982). Indeed, three years earlier the inaugural issue of the *European Journal of Science Education* had included a paper that found gender differences in attitudes towards the sciences in U.K. school students (comparing attitudes to physics and biology) and argued that these were being used to express sex differences (Ormerod, 1979). Thus, the interest and achievement gap between genders in physics had already become a focus of concern in Britain, with the comparison with women's better performances in biology indicating that this was not due to a difficulty with scientific process. Attention ranged from school students (where researchers thought differences could be traced to early socialization rather than subject oppositions in the classroom) through to women at the university. Researchers combined statistical studies with qualitative, interview-based studies and sought to test diverse explanations, in particular focusing on psychological stereotyping. Writing on women university students, Ian Lewis argued that "the hidden curriculum of option choices and transmitted messages of stereotyping, in relation to the study of physics, do play a critical part in reducing the numbers of potentially capable students unnecessarily" (Johnson and Murphy, 1984; Lewis, 1983, p. 193; and 1984). In 1987, the Manchester sociologist Alison Kelly edited a special issue with seven articles and an annotated bibliography on Science and Gender for the recently renamed *International Journal of Science Education*. Its introduction makes sobering reading, both because the situation has barely changed in regard to the central question of women's participation in physics and because as we shall see the essential intellectual achievements Kelly notes are now being advanced again, although in a more robust form. In particular, Kelly observed in 1987 that gender had been a concern of science educators for some time, and that over the past decade from an initial focus on underachievement in science tests (which risked blaming the victim), researchers had begun to engage a broader range of approaches, even beginning to ask not only how women were failing science, but "What is wrong with science that girls don't like it?" with research addressing the classroom interaction, teacher attitudes, and curriculum content (Editor's introduction: Gender and science, 1987, p. 259). The comparative breadth of the issue's contributions would also strike readers now, with discussion of education in developed and developing countries ranging from indigenous technology in Sierra Leone to computing, as well as two papers offering historical approaches. Kelly noted that reports showed that girls in Kuwait and Thailand achieved as well or better than boys in physical science, and that Beatriz Ruivo had argued that the relative lack of commercial significance of the subject in less industrial nations might have opened participation in places like Portugal to less powerful social groups (Editor's introduction: Gender and science, pp. 260–261). Noting the relative strength of women's participation in Eastern European and then communist countries, we can also see that researchers were unlikely to see conditions in the West as universal.

Twenty-nine years later, *Physics Education Research* editors offered a similar lesson through the example of Türkiye (with women there reaching 53% participation in university level physics in 2010) and actively sought international engagement, receiving half of the 42 proposals and six of the final 17 papers published by scholars outside the U.S. They had not expected extensive literature review papers, but the three that they published offered guiding perspectives on sociocognitive elements of undergraduate education and on sociopsychological dimensions of becoming a member of a physics community, and synthesized the history of Physics Education Research with an account of current methods (Kelly, 2016; Lewis *et al.*, 2016; and Traxler *et al.*, 2016). Collectively, these papers address the range of subtly mixed social and intellectual factors important in addressing deep-seated cultural phenomena, and also speak to researchers' needs to bring order to what they understood as a rapidly emerging field. Gender, once a minor concern, is now clearly being integrated in work throughout the field, thereby potentially addressing the discipline much more fully than previously. A measure both of the topic's previous standing and the new impulse towards more encompassing research is provided by an editorial note in which Rachel Scherr showed that only 7% of the more than 400 papers published in *Physics Education Research* since its first issue in 2005 had been devoted to gender, and that 80% of those were focused on performance gaps (Scherr, 2016).

Angela Kelly's synthesis of theoretical research and empirical work devoted to improving women's experience of undergraduate courses addresses theories of intelligence and creating a classroom climate, the benefit of positive familial role models and classroom models in creating a supportive environment, and curricular and institutional support. The approach integrates social and cultural features in ways now beginning to be thought to be empirically effective (Kelly, 2016). Similarly, Lewis and colleagues' account of the factors found relevant to increasing a social and cultural sense of belonging in physics has immediate practical goals, concluding with suggestions for educators (Lewis *et al.*, 2016). Developing a more encompassing and interpretive theoretical approach, Traxler *et al.* (2016) engage a historical perspective on recent key lines of research in offering a valuable account of research directions and arguing for an engagement with more performative and open-ended approaches to gender. They review studies of standardized measures, such as, in particular, the Force Concept Inventory, which has been investigated more thoroughly than any other. As they note, the important 10-year study of Doktor and Heller (2008) has established a persistent gender gap averaging over 15% in pretest scores and over 13% in posttest scores (and another study has found gaps despite no gap in course grades). A gender-theoretic reading of the test has also picked out the extent to which the test questions display stereotypically masculine contexts such as hockey and cannonball. Laura McCollough produced a revised version of the test that substitutes stereotypically female questions, and these obtained different results—but at the expense of male performance decreasing (McCullough, 2004). This is one area in which recent research has introduced important new arguments, with Traxler and colleagues showing in two valuable papers that the Force Concept Inventory featured questions that were structurally unfair, and then using a range of different tests to partition the gender gap into different components. In the first paper published in 2018, they analyzed three samples looking for gender asymmetries using classical test theory, item response theory, and differential item functioning. These methods highlighted six items that appeared substantially unfair to women and two items biased in favor of women, leading to the recommendation to report test results with and without the items (Traxler *et al.*, 2018). The following year, Traxler worked with Henderson and Stewart to construct modified conceptual inventories eliminating invalid or unfair items. They used hierarchical linear regression (HLR) to analyze the gender gaps controlling for academic performance (which was measured by test average or ACT/SAT math percentile), and prior physics preparation measured by pretest scores. The gender gap could then be "partitioned" to determine which factors were most important to the observed gender differences and whether the relative importance of the factors was consistent across instruments and institutions (Henderson *et al.*, 2019).

One of the most important analytic steps that (Traxler *et al.*, 2016) develop is a critique of treatments of gender as a binary explanatory model of factors that may influence student conceptual performance and attitudes, responses to new curricula, classroom achievement and retention in physics, often implicitly assuming that male characteristics provide the template for success. Just as research on learning has profitably moved from deficit-based studies of misconceptions to exploring the construct of "conceptual understanding," Traxler and colleagues build on the work of Judith Butler, and many studies of gender in research on education, to advocate analyzing gender as a performance rather than a fixed binary, thereby opening research up to the analysis of individual experience and the interrelations between different elements of personal identity, considering also the intersection of race and gender (Traxler *et al.*, 2016, pp. 5–7, 7–9). Similarly, Gonsalves *et al.* (2016) shift from focusing on concept acquisition to identity construction (and identity trajectories, in particular), drawing on masculinity studies to provide an analytical framework for the examination of respects in which learning physics through laboratory practices involves identifying with specific physics practices and potential careers as much as it does the acquisition of concepts. Three case studies provide diverse and nuanced accounts of the construction of masculinity (even when gender is not explicitly part of the narrative of all participants) associated with specific technical and analytical skills that are also integral to performing as a competent physicist—and as importantly, to being recognized as one. Their treatment of subtly gendered activities enables physics education researchers to position practices in physics within the social forces that shape our understanding of them. Similarly, significant work has been undertaken from the perspective of feminist standpoint theory, examining the lived experiences of womens' lives in physics and astronomy and building knowledge from this collective social consciousness—without including, interviewing, or including men as a comparison, and, in particular, showing that women astronomers interviewed defined success in terms of work-life balance and long-term life goals (Barthelemy *et al.*, 2015, 2016).

I have noted that physics education research has too often been pursued without a clear awareness of studies in other relevant but more broadly based literature. Remedying this lack by investigating the treatment of gender across the sciences, Sax *et al.* (2016) have valuably offered a historical study comparing U.S. women physics majors with all other women and women students of other STEM fields over four decades. Their analysis shows that a distinctive profile of the average female physics student has remained largely consistent: "Women who intend to major in physics tend to be confident in their math abilities, value college as an opportunity to learn, plan to attend graduate school, and desire to make theoretical contributions to science" (p. 5). Yet the number of students filling this profile, and interest amongst women in physics is extremely low (between 0.1 and 0.2%) and declining relative to interest in other STEM fields such as biology and engineering. Drawing on Whitten's (2012) argument for a feminist approach and analysis of the ways physics might change, they conclude "If the field of physics wishes to attract more women, it may be necessary to change the perception of the field and the focus of the major in order to reach a broader audience" (Sax *et al.*, 2016, p. 5).

6.4 COMBATTING BIAS: RACE

Education has been critical for minority communities, offering a pathway to social improvement and status while also remaining an ambivalent symbol of the power of cultural elites, and science has undoubtedly played an ambiguous role in these kinds of tensions. Science has usually been treated as one of the crowning intellectual and cultural achievements of civilization, regarded as a key to power but as pure in itself, although work in the sciences often served imperial and national geopolitical goals as well as straightforward military purposes both in the expansion of Western power in the colonial period and through the Cold War. Perhaps it should be no surprise that scientific education can raise complex double-binds for members of minority communities. An example of these tensions was played out in the early twentieth century in an important debate between Booker T. Washington, who argued for practical "industrial" education that would prepare African Americans for the workforce in Southern states, and W.E.B. DuBois, who argued for a college- and university-based education that would unlock the cultural dreams of knowledge. Elmer S. Imes earned the U.S.'s second African-American Ph.D. in physics for experimental studies of molecular spectra at the University of Michigan in 1919 and followed the example that DuBois had articulated—something that may have helped set a significant institutional example. Living in New York, Imes experienced the Harlem renaissance and in 1930 took up a professorship at Fisk University, where he taught a course on Cultural Physics that Ronald Mickens has described as emphasizing the intermingling of physics with other disciplines and underlining a cultural value equal to that of the humanities and the classics (Mickens, 2018, p. 33). Nearly one hundred years later, physics education researchers have found that historically black colleges and universities are amongst the few that have focused on recognizing talent and interest amongst students otherwise unprepared for physics; they have also been particularly significant for the relatively few black women who have gained baccalaureates or higher degrees in physics (Leggon and Willie Pearson, 1997; and Whitten *et al.*, 2003a, 2003b).

We need to look outside core disciplinary journals in physics education to find the most significant research on race, and this has often been combined with studies of gender. In the course of its first year of publication in 1994, the *Journal of Women and Minorities in Science and Engineering* published three articles that illustrate the character and scope of this research, ranging across all educational levels. Working at Miami University in Ohio, Judith Kahle and Arta Damnjanovic used a survey methodology to examine the influence of inquiry-based activities in teaching electricity to elementary school students across gender and race, finding this improved perception for all but also discerning differences in Caucasian and African-American students (Kahle and Damnjanovic, 1994). Barbara Bruschi and Bernice Anderson at the Educational Testing Service and the National Science Foundation examined average proficiency scores in four science content areas and at three different ages in National Assessment of Education Progress (NAEP) science proficiency data for 1990. They found differences between and among the three races considered, noting that the advantages of white male students widened over time, the gap with African-Americans was the greater than that with Hispanics, and that females were favored in the nature of science studies over all ages (Bruschi and Anderson, 1994). Daniel Solorzano at UCLA conducted the first U.S. national study of the Chicano and Chicana doctorate population across physical, life and engineering sciences, finding them the most underrepresented of the three major ethnic groups and (in a telling indication of some of the effects of what has become known as the leaky pipeline) noting the significance of small, mostly private and primarily teaching undergraduate institutions in the backgrounds of chicana/os students who subsequently went on to research degrees (Solorzano, 1994).

In physics education research, in particular, some of the most significant early research on ethnic minorities was developed by Maria Ong in NSF-funded longitudinal studies of minority female students negotiating incongruities between their field of study, ethnicity, and gender—attempting, with more or less success, to embody the identities of ordinary women, ordinary persons of color, and ordinary aspiring scientists. Using interviews with persevering students that began in 1996 to examine the conditions of success, Ong's (2005) analysis centered on the "body-projects that these women undertook responding to the institutional environment and perceptions of their peers to surmount barriers that race, ethnicity and gender posed in their local physics communities over a period of eight years." Ong showed that women pursued diverse strategies: of fragmentation, that includes gendered passing and racial passing in a temporary splitting oneself to minimize cultural differences between oneself and a community; and of multiplicity, in which one pursues a less stable but more wholistic occupation of several, sometimes competing identities, such as stereotype manipulation and performances of superiority. Although responding to exclusionary tactics in these ways often has significant personal costs, their participation also offers minority physicists insights into the limitations of customary perspectives. The astronomer Chandra Prescod-Weinstein has recently built on feminist standpoint theory to develop an epistemological analysis highlighting physicists' different responses to String Theory (despite its lack of empirical foundation) and the perspectives of women of color, arguing that white supremacist racial prestige asymmetries have produced an antiempiricist epistemic practice among physicists, that she calls "white empiricism" (Prescod-Weinstein, 2020).

Prescod-Weinstein's major (2021) book *The Disordered Cosmos: A Journey into Dark Matter, Spacetime, and Dreams Deferred* joins Hakeem Oluseyi's similarly thoughtful autobiography *A Quantum Life: My Unlikely Journey from the Street to the Stars* (2021) in offering gripping accounts that highlight the contingencies that minority scientists have negotiated (which will typically seem foreign to the white physicists with whom they have otherwise shared careers), in building their lives in physics. Lived accounts of this nature should become central to academic research aiming to incorporate minority perspectives for the future of physics, read alongside the work that has recently begun to explore the relations between individual identity and social constructs (Hyater-Adams *et al.*, 2018, 2019). Amongst such studies (Avraamidou, 2020) is unusual for its exploration of the experience of a woman, trained in Türkiye and the U.S., now working in Western Europe, who has experienced the intersection of gender, minority, immigrant and religious identities (sometimes in contest), with multiple barriers that positioned her as Other and hindered her sense of belonging, despite her self-identity as a science person. Supporting Avraamidou's argument for the conjunction rather than isolation of identities, and for research examining the politics of recognition in diverse cultural and geographical contexts (developed further in Avraamidou, 2021), we can note that migrant experience is significant for many physics researchers in particular, and for the last twenty years, a majority of U.S. doctoral students in physics have been foreign citizens.

Rodriguez *et al.* (2022) have recently undertaken the valuable service of providing a historical overview of critical race theory and feminist standpoint theories (considering also the related but independent field of intersectional studies). They introduce researchers to the conceptual foundations of these fields, and address common critiques in the light of continued controversy over their contributions. Critical race theory has been developed in physics education research following work on the educational pathways of black women physicists (Rosa and Mensah, 2016), and is underpinned by four principal tenets: that racism is ordinary (though often invisible to elites), that rights will only advance when they align with the interest of the elites (an insight that highlights the practical difficulties of achieving change), that race is socially constructed from the products of social relations, and that unique voices of colour must be heard to understand racism. The approach meshes well with the feminist standpoint theory, building on Marxism to create knowledge and social understanding from the perspective of the oppressed and developed in understandings of scientific work (and physics) by Sandra Harding, Evelyn Fox Keller, Barbara Whitten and others. Drawing together work that has sometimes been pursued first under other frameworks, Rodriquez and colleagues highlight both more longstanding qualitative and recent quantitative studies beginning to incorporate these frameworks within physics education research (Van Dusen and Nissen, 2020; and Nissen *et al.*, 2021). Their study also offers an excellent basis for further research, and many studies also offer important insights both on the coping strategies individuals have developed, and for institutional change (Johnson *et al.*, 2017; and Dickens *et al.*, 2020). In sum, research recognizing the human diversity of engagement in physics—and intersectional research, in particular—is absolutely necessary to understand when minoritized groups experience inclusion or exclusion, and how and why this occurs; this could well have policy implications as well as further research to improve teaching, learning and physics curricula (Avraamidou, 2020, p. 338).

6.5 THE COLD WAR AND THE FOUNDATIONS OF THE MODERN PHYSICS DISCIPLINE: PEDAGOGICAL HISTORIES AND THE ANTHROPOLOGY OF PHYSICS

Although it is highly important to meet broader cultural and social challenges by addressing diversity and inclusion, the longstanding persistence of the specific demographic profile of physics despite significant social change should alert educators to the need to cultivate long-term structural and historical perspectives to address these issues successfully. This is essential to develop an adequate understanding of the institutional and social character of the physics discipline, in particular, but also discloses the historically contingent nature of the primary cultural and ethical issues that physics has seemed to pose at different times. There are at least three ways in which developing a long-term perspective will deepen physics education researchers' treatment of physics as a human endeavor, combining the methods and insights of historical and anthropological research to deepen the analysis of the social specificities that have marked the development of physics over time—and thereby helping to prepare more thoroughly for future educational needs.

The first lies in cultivating a multifaceted causal understanding of the interrelations between intellectual, material and institutional developments generally. There are many historical resources to guide this endeavor; for overviews see Richard Staley's (2013) account of the history and historiography of the physical sciences and Helge Kragh's (1999) book length study of the physics discipline, while Agar (2012) provides an integrative perspective on physics within the complex of the sciences as a whole. Perhaps the most important overarching point to note is the extent to which the forms of physics adopted after World War II were strongly shaped by the material and cultural power that accrued to physicists as a result of their provision of the weapons of war—as well as those that might be required for future battle and performance in the ideological hot-house of the Cold War that ensued between communist and Western nations vying for geopolitical power. Peter Galison's magisterial account of particle physics is particularly pertinent for providing a general understanding of the modern physics discipline and the varied choices made as physicists came back from war and sought new independence while utilizing the funds and prestige associated with research relevant to (often quite general) military aims (Galison, 1997). His study investigates the relations between two distinct but interrelated strategies developed in detector physics, described in his title as *Image and Logic* traditions. These focused initially on photographs of singular events and on statistical treatments, respectively, and were subsequently melded in the electronically controlled detection systems now in use. In addition to his treatment of laboratory leaders, instrument makers, detector technologies and image readers, Galison has devoted significant attention to the distinctive social organization of high energy physics and, in particular, the diverse practices that different groups at CERN and elsewhere have developed to achieve statistical significance levels of 5σ[1] and manage the work of writing collectively (Galison, 2003). Author lists on projects that often take decades to reach fruition have numbered in the thousands, as they did, for example, in the detection of the Higgs Boson in 2012, and the physics community should also be recognized for its development of genuinely novel sociological forms of knowledge production. Yet, educators' awareness of high energy physics should be complemented by recognizing the diversity of practices in different fields of physics—always more numerically and commercially important than particle physics—but also the extent to which they are interrelated and produced in a competitive environment. Historical work on other fields such as the culturally important fields of cosmology and astrophysics and the commercially important field of condensed matter physics is essential for this; Joseph Martin's accounts of the institutional politics and prestige asymmetries that have marked the development of condensed matter physics are particularly valuable (Martin and Janssen, 2015; and Martin, 2018).

The second respect in which long-term perspectives can deepen physics education research concerns historical reflection on pedagogy in particular. No one has done more to integrate both institutional and intellectual histories of physics since World War II with education than David Kaiser. His scholarship has been particularly important in offering pedagogical histories for theoretical developments, developing accounts that range from reconstructing how Feynman diagrams were taken up in subtly different ways dependent on the specific resources with which physicists approached them, to considering the significance of the collapse of funding for particle physics in 1971 for promoting new disciplinary prospects—and enabling the combination of cosmological and particle perspectives by students who sat in classes in general relativity and particle physics and recognized an intellectual unity that was inaccessible to their teachers (Kaiser, 1998, 2005b, 2005c, 2006a, 2007). Among the significant scholarship on pedagogy in Cold War America, Kaiser's studies of the anti-Communist campaigns of the 1950s are particularly pertinent to nature of science concerns with the ethical and social values informing physics. Robert Oppenheimer's experience provides one possible entry. Biographies from David Cassidy (2005); Bird and Sherwin (2006); and Thorpe (2006) disclose the effects of changing national priorities on his career, and Hegeman (2017) uses Oppenheimer as a key example of the tensions between academic freedom, self-governance, and State interests. She also examines the changing meanings of academic freedom and academic labor in the transition from the Cold War to the neoliberal university, determined more by neoliberal common sense oriented around private interests. While these examples illuminate physicists engagement in the cultural environment of their period, Kaiser's account of Anthony Chew's "democratic bootstrapping" anti-fundamentalist versions of "nuclear democracy" takes a further step in their mutual interplay, showing how one American physicist incorporated social values in his development of nuclear theory, and the social life of a

---

[1] That is, the probability of achieving such a result were there not really anything going on, is five standard deviations of the normal distribution, corresponding to about a one-in-a-three-and-a-half-million chance that the findings are just a result of random variations.

research group that was defined as much through an exclusionary approach to disciplinary alternatives as by suspending formal hierarchy within the group (Kaiser, 2002, 2005a, 2006a).

In 2005, Kaiser published a highly valuable edited volume drawing together examples of pedagogy in the nineteenth and twentieth centuries, examining the practices that have produced scientists and engineers in specific cases in Europe, Japan and the United States. The conclusion to the volume, written with Andrew Warwick, argued for a melding of Thomas Kuhn's focus on textbooks in the propagation of normal science with Michel Foucault's argument that power in social relations does not inhibit or conceal knowledge, but is central to its production, which they combined with Foucault's attention to the way that the minutiae of everyday practices have the possibility of enabling new capacities in order to generate accounts of significant historical change in the sciences (Warwick, 2003; and Warwick and Kaiser, 2005). In 2013, Massimiliano Badino and Jaume Navarro noted tensions in this attempt to articulate the relations between knowledge and power. By drawing together a collection of accounts of textbooks written in the dynamic "revolutionary" period in which quantum mechanics was being formed, they sought to counter Kuhn's insistence that as exemplars of normal science, textbooks offer conceptually and historically misleading perspectives on the making of science—or, in terms more familiar in physics education research, on the nature of science—because their commitment to current interpretive approaches often leads them to rewrite the history of science from the perspective of the present (Badino and Navarro, 2013). Both the many examples of physics and engineering pedagogy that Kaiser and Navarro and Badino have assembled, and their methodological and theoretical reflections should help researchers in physics education investigating current educational resources.

While addressing significant questions in the relations between education and research, case studies of this kind remain limited by their particularity of focus. Two powerful examples of works that address educational lifeworlds more comprehensively provide a third respect in which physics education research can benefit from long-term structural perspectives. This rests on the extent to which they can illuminate the subtle cultural dimensions of education in physics. The first example is Warwick's (2003) study of training in mathematics at the University of Cambridge in the nineteenth and early twentieth century, which offers an exhaustive, archivally-based treatment of the intellectual and physical regimes of training in Cambridge. This typically complemented worked practice under the direction of mathematical coaches with college routines and sport—and Warwick's account of this cultural elite both offers insight into its gendered masculine rendering and shows how specific technical practices shaped mathematicians' interpretations of electron theory and relativity. Three chapters "Writing a pedagogical history of mathematics" and developing a historical ethnography of educational practice (studying both the way student bodies were exercised and the work of mathematical coaches), are particularly pertinent for their combination of methodological and historical insights into the pedagogical process (Warwick, 2003, Chap. 1, 4, and 5). My second example has already proved important for physics education researchers. Often cited for its argument that physics has usually been pursued as if it has a culture of no culture, Sharon Traweek's *Beamtimes and Lifetimes* (1988) offers an extensive comparative anthropological study of high energy physics at the Stanford Linear Accelerator and at the Japanese facility KEK based on work there in the 1980s. She analyzes the implicit messages conveyed by text boxes and images included in physics textbooks to show how physicists presented their work in gendered terms as offering a spearhead into the penetration of nature; displays the different modes of presentation of experimentalists focused on particle accelerators as machines generating data, and of theorists who treat them more transparently as instruments simply recording nature; and offers a brilliant account of cultural differences in the pursuit of physics in the U.S. and Japanese facilities.

Traweek's ethnographic work examines the intersection of identities within dominant cultures, and shows that the characteristics that were understood to favor progress in the highly competitive environment of physics groups in the U.S. were very different from those cultivated in Japan. Japanese physicists placed great responsibility for the future on young physicists, but also an obligation to their teachers: training the next generation was critical to become a first-rate particle physics nation in accord with the tenets of *amae*, promoting interdependence, especially across generations, in a mutual responsibility and obligation that is a crucial value in Japan. In contrast, Americans believed that each individual physicist must provide the best physics possible, consistent with American *individualism* on the model of laissez-faire economics, in which an unfettered marketplace of ideas selects the best contributions, with self-interests necessarily and properly competing. In the 1980s, Traweek showed that while the style in the U.S. was informal, the group structure was strongly hierarchical, with decisions made by the group leader subsequently imparted to group members to be implemented. In contrast, Japanese values directed leaders to consult with group members, requiring them to advise thoughtfully, and group members would cooperate with a decision if they believed this process had been respected. Similarly, no strict divisions of labor marked Japanese groups as they did American groups (Traweek, 1988, ch. 5, esp. pp. 145–152). Although developing Weberian ideal-type characterization and recognizing that both patterns exist in both places and no person fits them entirely, Traweek found strongly marked distinctions. Her chapter on "Pilgrim's progress: male tales told during a life in physics" explores the subtly different perspectives dominant in different periods of training and work in physics, from undergraduates learning approximations to the truth while the margins feature images of (male) scientist-heroes, to postdoctoral researchers caught in a double bind between official descriptions of group work as cooperative and the message that only competition and (careful) forms of insubordination will ensure progress. There is a gender corollary:

> These stories about a life in physics define virtue as independence in defining goals, deliberate and shrewd cultivation of varied experience, and fierce competition with peers in the race for discoveries. Independence, experience, competition, and individual victories are strongly associated with male socialization in our culture. By contrast, recent studies in Japan suggest that these are the qualities associated with professionally active women, not men. Women are seen as not sufficiently schooled in the masculine virtues of interdependence, in the effective organization of teamwork and camaraderie, commitment to working in one team in order to complete a complex task successfully and consulting with group members in decision making, and the capacity to nurture the new group members in developing these skills.

Traweek goes on to note that while the virtues associated with success vary dramatically across these two cultures, “the virtues of success, whatever their content, are associated with men” (Traweek, 1988, p. 104). This empirically founded, carefully open treatment of both the characteristics associated with success and those associated with gender as socially constructed can be read historically for the understanding it provides of conditions pertaining in a particular period. It helps explain why focusing on particle *theory* and not on the *practices through which it was taught and developed further*, physicists could believe their discipline to be independent of social values in one moment, while inscribing them unconsciously in another—and in still other moments sometimes consciously celebrating their work in terms of the social values that it displays. Methodologically, Traweek’s research does still more than this, conveying the analytic tools to move between explicit and implicit messages and recognize tensions between them; a facility that at least potentially opens the possibility of change—and this can be informed by a historical understanding of the multidimensional accounts of pedagogical change that are offered by Warwick, Badino, Navarro, Kaiser and others. These historical and anthropological studies offer physics education researchers (and perhaps physicists themselves) case studies for classrooms and seminar discussions, but they also develop resources complementary to more contemporary treatments of gender, disciplinary and national identities, opening significant perspectives on the structural nature of the issues concerned. Bøe *et al.* (2018) offer an account of the implied student who can help teachers incorporate historical and sociocultural dimensions of physics in more traditional classroom contexts.

6.6 CLIMATE CHANGE AND THE FUTURE PHYSICS CLASSROOM: MEETING THE ANTHROPOCENE

Physics education should surely meet the most significant challenges of the physics discipline and our time: this final section examines the question of human endeavor in the broadest and most general framework now offered for understanding human work with the earth. Paul Crutzen is now better known for offering a name for a potential new geological epoch than for his Nobel Prize-winning work on the formation and decomposition of atmospheric ozone. Geologists distinguish the current epoch, the Holocene, by the geological record of the end of the last Glacial Period about 9700 years BCE. In the middle of a 2000 conference, Crutzen called out to stop a speaker referring to the Holocene, for the term suddenly seemed wrong given how much had changed. “Stop it!” he said, “We are no longer in the Holocene, we are in the Anthropocene,” subsequently publishing a groundbreaking newsletter comment with Eugene Stoermer, the marine biologist whose term he had borrowed. This raised many of the issues that have helped “the Anthropocene” become a metaphor, at the same time that its naming posed for geologists the formal question of whether to recognise a new epoch (Crutzen and Stoermer, 2000). The first point concerned the interacting scales at which change had occurred. Crutzen and Stoermer identified a range of impacts resulting from human and cattle population explosions: fossil fuel depletion, SO_2, synthetic N, species extinctions, greenhouse gases, toxic substances, chlorofluorocarbon gases, loss of mangroves, and depletion of fisheries. They wrote, “Considering these and many other major and still growing impacts of human activities on earth and atmosphere, and at all, including global scales, it seems to us more than appropriate to emphasize the central role of mankind in geology and ecology by proposing to use the term ‘anthropocene’ for the current geological epoch” (Crutzen and Stoermer, 2000, p. 17). The second issue they highlighted was the question of dating the period’s onset. Pointing to Watt’s steam engine, they proposed the late eighteenth century because it was marked by the first rise in greenhouse gas emissions, and “the global effects of human activities have become clearly noticeable.” While they suggested dating was somewhat arbitrary, much discussion has been stimulated by proposals offering diverse perspectives on responsibility by tying the name of the period to different arguments about its causes. Was it early humans first changing carbon dioxide and methane concentrations with agriculture and rice paddies more extensive than previously thought, as Bill Ruddiman first proposed (Ruddiman, 2003; and Ruddiman *et al.*, 2020)? Or should it be tied to European colonization of the Americas, with the Orbis spike of 1610 registering a sharp decrease in carbon dioxide concentrations that occurred as a result of forest regrowth after the large-scale murder of indigenous peoples with American colonization (Lewis and Maslin, 2015)? Crutzen and Stoermer had linked the date with early industrialization and global warming, writing of “mankind,” but many commentators would bring this home more specifically to capitalism, oil firms, and government energy policies. Or should we instead recognize the “great acceleration” following World War II, when production, change and waste accelerated in all sorts of terms (Steffen *et al.*, 2011)? This is the period that geologists have settled on so far in their lengthy decision-making process, for the magnitude of change and clear geological markers available (Zalasiewicz *et al.*, 2021). Yet all of these dates have been contested, in part because the question of responsibility carves humanity at the joints in so many more pointed ways than knowing simply that the effects of a small number of individuals, firms and nations are changing the earth we all inhabit, and that all of us can work to improve its fate. Thus, while thinking in terms of the general category of mankind is certainly valuable, as Bonneuil and Fressoz (2016) show, understanding the very different historical responsibilities and current carbon footprints of diverse groups is certainly more important.

We can sharpen the disciplinary and ethical questions this engages first by recalling that I noted above that the Next Generation Science Standards explicitly raise ethical questions only when considering the earth and space sciences; and secondly by recognizing that it is important to approach the physical sciences widely enough to encompass their engagement both in changing the earth itself, and in developing the observational tools required to characterize the nature of that change through geophysics and the climate sciences. Scientists have increasingly recognized the extent to which the earth and climate sciences have been shaped by commercial and colonial development, with considerable work required to strengthen data coverage in areas less strategically relevant to commercial and state interests. Indeed, Brönnimann and Wintzer (2019) use a study of data coverage maps to show that data depend on political, economic and technological factors and argue that this context should be seen as a source of information to be communicated along with the data. Our understanding of "global" sciences should be as attentive to geopolitical difference as our concern for physics as a human endeavor can be to diverse identities. Often the tools required to measure changes in earth and climate have also been important in enabling the activities that have changed the earth—such as satellites monitoring weather and climate taking a place within increasingly wide-spread and fine-grained military and communication networks. The distributed, complexly interrelated nature of the phenomena and the diverse, often competing interests engaged in technological development have posed distinctive problems both for the articulation and for the widespread public acceptance of robust knowledge about the climate. Here historical work offers significant guidance in how to explore the unusual implications these have for understanding the nature of science by charting the ways that the 150 year emergence of our understanding of climate change reflects such highly temporally and spatially distributed phenomena that it has required a particularly challenging set of interdisciplinary relationships—at odds with previous norms in scientific research (Weart, 2003). The development of climate modeling from meteorology in the 1970s and later of integrated climate and earth system models must also be understood intellectually, materially, and socially, as Paul Edwards, Matthias Heymann, Martin Mahony and colleagues have shown by documenting the emergence of climate modeling and examining the diverse ways it reflects colonial heritages and tensions between global North and South nations (Edwards, 2010; Heymann, 2010; Heymann and Kragh, 2010; Mahony and Endfield, 2018; Heymann and Dalmedico, 2019; and Miguel *et al.*, 2019).

These long-term structural relations embody significant political relationships closely implicated with commercial and state power. They complement, and should be engaged as seriously as climate science skeptics' arguments against the scientific consensus on climate change—which have so far received far more attention. Nancy Oreskes and Erik Conway's vital study of climate science denial shows that this has rested upon industrial funding and the political organization of several key politically engaged scientists who were almost always trained first in the physical sciences, and who were committed to free-market principles (Oreskes and Conway, 2010). Their work first stacking government committees, then attempting to weaken public trust in the findings of the Intergovernmental Panel on Climate Change, and still more recently promoting energy industry interests should be understood in company with Joshua Howe's (2014) decade by decade study of climate scientists' gradual engagement with political organization on a national and international scale. Howe shows that climate scientists followed the model of the earlier environmental movement and legislation, but argues that they were tragically committed to the belief that putting the climate sciences on a firm scientific footing first would compel the appropriate action. Our students need to understand how much more is required than that. This will mean engaging deliberately with what otherwise risks remaining simply the unconscious politics of a cultural elite. Physics education research has done a great deal to help physicists realize greater gender and racial diversity in their classrooms, but responding appropriately to climate change will require cultivating both greater disciplinary diversity (teaching methods and content new to many physicists), and a readiness to meet technical tools with appropriately sensitive political engagement. Physics is better understood as a human endeavor.

REFERENCES

Agar, J., *Science in the Twentieth Century and Beyond* (Polity, Cambridge, 2012).

Aikenhead, G. *et al.*, *Views on Science–Technology–Society (From cdn.Mc.5)* (Department of Curriculum Studies, University of Saskatchewan, Saskatoon, 1989).

Australian Curriculum, A. a. R. A., see https://www.australiancurriculum.edu.au/senior-secondary-curriculum/science/physics/structure-of-physics/ for senior secondary curriculum, physics, structure of physics. Australian curriculum (2022).

Avraamidou, L., J. Res. Sci. Teach. **57**, 311–341 (2020).

Avraamidou, L., J. Res. Sci. Teach. **59**, 58–94 (2021).

Badino, M. and Navarro, J., *Research and Pedagogy: A History of Quantum Physics Through Its Textbooks* (Edition Open Access, 2013).

Barinaga, M., Science **263**(5152), 1468–1472 (1994).

Barthelemy, R. S. *et al.*, Int. J. Gender Sci. Technol. **7**(1), 57–73 (2015).

Barthelemy, R. S. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020119 (2016).

Beichner, R. J., *Getting Started in Physics Education Research*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2009).

Bell, R. L., Ph.D. thesis (Oregon State University, 1999).

Bird, K. and Sherwin, M. J., *American Prometheus: The Triumph and Tragedy of J. Robert Oppenheimer* (Alfred A. Knopf, New York, 2006).

Bøe, M. V. *et al.*, Sci. Educ. **102**, 649–667 (2018).

Bonneuil, C. and Fressoz, J.-B., *The Shock of the Anthropocene: The Earth, History and us (D. Fernbach, Trans.)* (Verso, London, 2016).

Brönnimann, S. and Wintzer, J., WIREs Clim. Change **10**(2), e559 (2019).

Bruschi, B. A. and Anderson, B. T., J. Women Minor. Sci. Eng. **1**(3), 221–236 (1994).

Cassidy, D. C., *J. Robert Oppenheimer and the American Century* (Pi Press, New York, 2005).

Cooley, W. and Klopfer, L., *Test on Understanding Science (Form w)* (Educational Testing Service, Princeton, NJ, 1961).

Crutzen, P. J. and Stoermer, E. F., IGBP [Int. Geosphere-Biosphere Programme] Newsletter **41**, 17–18 (2000).

Dickens, D. *et al.*, Phys. Teach. **58**(5), 335–337 (2020).

Docktor, J., and Heller, K., Paper Presented at the AIP Conference Proceedings (AIP, 2008), p. 1064.

Editor's introduction: Gender and science, Int. J. Sci. Educ. **9**(3), 259–261 (1987).

Edwards, P. N., *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (MIT Press, Cambridge, MA, 2010).

Etkina, E. *et al.*, Am. J. Phys. **67**(9), 810–818 (1999).

Fish, D. *et al.*, *Perimeter Inspirations Teacher's Guide* (Perimeter Institute, Warterloo, ON, 2013).

Galili, I., Sci. Educ. **28**(3–5), 503–537 (2019).

Galison, P., *Image and Logic: The Material Culture of Microphysics* (Univ. of Chicago Press, Chicago, 1997).

Galison, P., *Scientific Authorship: Credit and Intellectual Property in Science*, edited by M. Biagioli and P. Galison (Routledge, New York/London, 2003), pp. 325–355.

Gonsalves, A. J. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020120 (2016).

Götschel, H., Sci. Technol. Stud. **24**(1), 66–80 (2011).

Hadzigeorgiou, Y., Phys. Educ. **41**(1), 42–46 (2006).

Hegeman, S., J. Acad. Freedom **8**, 1–15 (2017).

Heilbron, J. L., *The Dilemmas of an Upright man: Max Planck and the Fortunes of German Science (With New Afterword Ed.)* (Harvard University Press, Cambridge, MA, 2000).

Heller, P. and Hollabaugh, M., Am. J. Phys. **60**(7), 637–644 (1992).

Henderson, R. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(1), 010131 (2019).

Hestenes, D. *et al.*, Phys. Teach. **30**, 141–158 (1992).

Heymann, M., Stud. History Philos. Mod. Phys. **41**(3), 193–200 (2010).

Heymann, M. and Dalmedico, A. D., J. Adv. Mod. Earth Syst. **11**(5), 1139–1152 (2019).

Heymann, M. and Kragh, H., *Modelling and Simulation in the Atmospheric and Climate Sciences* (Elsevier, 2010), Vol. 41(3).

Howe, J. P., *Behind the Curve: Science and the Politics of Global Warming* (Washington University Press, Seattle/London, 2014).

Hyater-Adams, S. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(1), 010132 (2018).

Hyater-Adams, S. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020115 (2019).

Ivie, R. *et al.*, *African Americans & Hispanics among Physics & Astronomy Faculty: Results From the 2012 Survey of Physics & Astronomy Degree-Granting Departments* (American Institute of Physics Statistical Research Center, College Park, MD, 2014).

Johnson, A. *et al.*, Phys. Teach. **55**, 356–360 (2017).

Johnson, S. and Murphy, P., Eur. J. Sci. Educ. **6**(4), 399–409 (1984).

Joint Physics Education, C., *Girls and Physics: A Report by the Joint Physics Education Committee of the Royal Society and the Institute of Physics* (Institute of Physics, London, 1982).

Kahle, J. B. and Damnjanovic, A., J. Women Minor. Sci. Eng. **1**(1), 17–28 (1994).

Kaiser, D., Stud. History Philos. Mod. Phys. **29**, 321–338 (1998).

Kaiser, D., Isis **93**, 229–268 (2002).

Kaiser, D., Representations **90**, 28–60 (2005a).

Kaiser, D., *Drawing Theories Apart: The Dispersion of Feynman Diagrams in Postwar Physics* (Univ. of Chicago Press, Chicago, 2005b).

Kaiser, D., *Pedagogy and the Practice of Science: Historical and Contemporary Perspectives* (MIT Press, Cambridge, MA, 2005c).

Kaiser, D., Soc. Res. **73**(4), 1225–1252 (2006a).

Kaiser, D., (2006b).

Kaiser, D., (2007), pp. 62–69.

Kapon, S. *et al.*, Sci. Educ. **102**(5), 1077–1106 (2018).

Kelly, A. M., Phys. Rev. Phys. Educ. Res. **12**(2), 020116 (2016).

Kragh, H., *Quantum Generations: A History of Physics in the Twentieth Century* (Princeton University Press, Princeton, 1999).

Lederman, N. G. *et al.*, J. Res. Sci. Teach. **39**(6), 497–521 (2002).

Lederman, N. G. *et al.*, *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. R. Matthews (Springer, Dordrecht, 2014), pp. 971–998.

Lederman, N. G. and O'Malley, M., Sci. Educ. **74**, 225–239 (1990).

Lederman, N. G. *et al.*, *The Nature of Science in Science Education: Rationales and Strategies*, edited by W. McComas (Kluwer Academic Publishers, Dordrecht, 1998), pp. 331–350.

Leggon, C. B. and Willie Pearson, J., J. Women Minor. Sci. Eng. **3**(4), 213–224 (1997).

Lewis, I., Eur. J. Sci. Educ. **5**(2), 185–193 (1983).

Lewis, I., *The Student Experience of Higher Education*, edited by I. Lewis (Routledge, London, 1984).

Lewis, K. L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020110 (2016).

Lewis, S. L. and Maslin, M. A., Nature **519**(7542), 171–180 (2015).

Mahony, M. and Endfield, G., Wiley Interdiscip. Rev.: Clim. Change **9**(2), 1–16 (2018).

Martin, J. D., *Solid State Insurrection: How the Science of Substance Made American Physics Matter* (University of Pittsburgh Press, Pittsburgh, 2018).

Martin, J. D. and Janssen, M., Hist. Stud. Nat. Sci. **45**(5), 631–640 (2015).

McCullough, L., J. Int. Women's Stud. **5**(4), 20–30 (2004).

McCullough, L., *Getting Started in Physics Education Research*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2018).

McDermott, L. C., *Niels Bohr Library & Archives, American Institute of Physics*, edited by D. Zierler (AIP, College Park, MD, 2020).

McDermott, L. C., *A View From Physics: Discipline-Based Education Research* (AIP Publishing, Melville, NY, 2021).

McDermott, L. C. *et al.*, J. Coll. Sci. Teach. **9**, 261–265, 135–140, 201–205 (1980).

McDermott, L. C. *et al.*, *Teaching Minority Students*, edited by I. J. H. Cones *et al.* (Jossey-Bass, San Francisco, 1983), pp. 59–72.

Megaw, W. J., Phys. Can. **48**(1), 25–28 (1992).

Meltzer, D. E. and Otero, V. K., Am. J. Phys. **83**(5), 447–458 (2015).

Mickens, R. E., Phys. Today **71**(10), 28–35 (2018).

Miguel, J. C. H. *et al.*, Sociologias **21**(51), 44–75 (2019).

Moché, D. L., Am. J. Phys. **44**(4), 390–391 (1976).

Mulvey, P. J. and Nicholson, S., *Physics Bachelor's Degrees: 2018. Results From the 2018 Survey of Enrollments and Degrees* (American Institute of Physics Statistical Research Center, College Park, MD, 2020).

Mulvey, P. J. *et al.*, *Trends in Physics PhDs: Results From the 2019 Survey of Enrollments and Degrees and the Degree Recipient Follow-up Survey for the Classes of 2017 and 2018* (American Institute of Physics Statistical Research Center, College Park, MD, 2021).

National Research Council, *Next Generation Science Standards: For States, by States* (NRC, 2013a).

National Research Council, *Next Generation Science Standards For States, By States* (NRC, 2013b).

Nissen, J. M. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(1), 010116 (2021).

Oluseyi, H. and Horwitz, J., *A Quantum Life: My Unlikely Journey From the Street to the Stars* (Ballantine Books, New York, 2021).

Ong, M., Soc. Probl. **52**(4), 593–617 (2005).

Oreskes, N. and Conway, E. M., *Merchants of Doubt: How A Handful of Scientists Obscured the Truth on Issues From Tobacco Smoke to Global Warming* (Bloomsbury Press, New York, 2010).

Ormerod, M. B., Eur. J. Sci. Educ. **1**(2), 177–190 (1979).

Park, W. *et al.*, Sci. Educ. **28**, 1055–1083 (2019).

Planck, M., Phys. Z. **10**, 62–75 (1909).

Porter, A. M. and Ivie, R., *Women in Physics and Astronomy* (American Institute of Physics Statistical Research Center, College Park, MD, 2019).

Prescod-Weinstein, C., Signs: J. Women Cult. Soc. **45**(2), 421–447 (2020).

Prescod-Weinstein, C., *The Disordered Cosmos: A Journey Into Dark Matter, Spacetime, and Dreams Deferred* (Bold Type Books, 2021).

Rifkin, M., Phys. Teach. **54**(2), 72–74 (2016).

Rodriguez, M. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 013101 (2022).

Rosa, K. and Mensah, F. M., Phys. Rev. Phys. Educ. Res. **12**(2), 020113 (2016).

Ruddiman, W. F., Clim. Change **61**, 261–293 (2003).

Ruddiman, W. F. *et al.*, Quat. Sci. Rev. **240**, 106386 (2020).

Rutherford, F. J. and Ahlgren, A., See https://ebookcentral.proquest.com/lib/cam/detail.action?docID=737311 for Science for all Americans (1991).

Sax, L. J. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020108 (2016).

Scherr, R., Phys. Rev. Phys. Educ. Res. **12**(2), 020003 (2016).
Solorzano, D. G., J. Women Minor. Sci. Eng. **1**(4), 253–272 (1994).
Staley, R., History Sci. **51**, 151–177 (2013).
Steffen, W. *et al.*, Philos. Trans.: Math., Phys. Eng. Sci. **369**(1938), 842–867 (2011).
Thorpe, C., *Oppenheimer: The Tragic Intellect* (University of Chicago Press, Chicago, 2006).
Traweek, S., *Beamtimes and Lifetimes: The World of High Energy Physicists* (Harvard University Press, Cambridge, MA, 1988).
Traxler, A. L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020114 (2016).
Traxler, A. L. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(1), 010103 (2018).
Trowbridge, D. E. and McDermott, L. C., Am. J. Phys. **48**(12), 1020–1028 (1980).
Trowbridge, D. E. and McDermott, L. C., Am. J. Phys. **49**(3), 242–253 (1981).
Van Dusen, B. and Nissen, J., Phys. Rev. Phys. Educ. Res. **16**(1), 010117 (2020).
Warwick, A., *Masters of Theory: Cambridge and the Rise of Mathematical Physics* (University of Chicago Press, Chicago, 2003).
Warwick, A. and Kaiser, D., *Pedagogy and the Practice of Science: Historical and Contemporary Perspectives*, edited by D. Kaiser (MIT Press, Cambridge, MA, 2005), pp. 393–409.
Weart, S. R., *The Discovery of Global Warming* (Harvard University Press, Cambridge, MA, 2003).
Whitten, B. L., J. Women Minor. Sci. Eng. **18**(2), 115–134 (2012).
Whitten, B. L. *et al.*, Phys. Today **56**(9), 46–51 (2003a).
Whitten, B. L. *et al.*, J. Women Minor. Sci. Eng. **9**, 239–258 (2003b).
Zalasiewicz, J. A. *et al.*, Earth's Future **9**(3), e2020EF001896 (2021).

CHAPTER

7

THE ROLE OF PHYSICS IN ACHIEVING SCIENTIFIC LITERACY IN THE PRESENT AND THE FUTURE

Hunkoog Jho

Jho, H., “The role of physics in achieving scientific literacy in the present and the future,” in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 7-1–7-22.

7.1 INTRODUCTION

Scientific literacy (hereafter SL) is one of the overarching goals in science education and in most countries, science and technology are taught as compulsory subjects in primary and secondary schools. In a lexical meaning, literacy refers to the ability to read and write (Oxford University Press, 2022) and more often refers to the particular ways of thinking about and doing reading and writing (Street, 2001). SL encompasses literacy in science matters and behavioral aspects in non-scientific situations. OECD (Organisation for Economic Co-operation and Development) defines this term as “an individual’s understanding of scientific concepts, phenomena and processes, and their ability to apply this knowledge to new and, at times, non-scientific situations” (OECD, 2019). In line with the rapid development of science and technology, the cultivation of science knowledge and scientific thinking has become important and curricular documents in many countries proclaim achieving scientific literacy as an important goal of education. For instance, the national science curriculum in the United Kingdom addresses the significance of scientific literacy as follows:

> Science has changed our lives and is vital to the world’s future prosperity, and all students should be taught essential aspects of the knowledge, methods, processes and uses of science. Through building up a body of key foundational knowledge and concepts, students should be encouraged to recognise the power of rational explanation and develop a sense of excitement and curiosity about natural phenomena. They should be encouraged to understand how science can be used to explain what is occurring, predict how things will behave, and analyze causes (Department for Education, 2015).

https://doi.org/10.1063/9780735425514_007

According to the national interest in SL, there are some efforts to evaluate students' SL and compare national performances. For example, PISA (Programme for International Student Assessment) is a triennial large-scale assessment for students' performances in relation to SL all over the world (PISA, 2017). It is used to present national competitiveness in STEM and Singapore, Macao, Estonia and Japan have been on the top of the list.

Moreover, Attitudes and beliefs are important to achieve SL. The development of science and technology has a great impact on our daily lives. The rise of socio-scientific issues such as fine particulate matter, radioactive effluent and genetically modified organisms requires us to make decisions with regard to the personal and social/regional aspects. Traditional characterization of SL is limited to enhance the ability of decision making because a human behavior is an unpredictable outcome from complex standards and intertwined with a variety of knowledge, emotion and experiences (Marchau *et al.*, 2019). Socio-scientific issues encompass scientific, societal, economic and ethical values and such of them are incompatible with each other (Ratcliffe and Grace, 2003; and Zeidler *et al.*, 2005). Thus, SL relevant to our lives requires individuals to take into consideration ethical/moral considerations besides scientific understanding, to perform cost-benefit analysis and to count cut-offs and trade-offs among conflicting evidence (Powell, 2021). This indicates that scientific understanding is not stressed as much compared to the traditional science classroom.

Physics is closely connected to the aforementioned elements of SL. Foremost, physics is a basic school subject and fundamental to the content knowledge of SL. STEM education, which is attracting attention today, is also relevant to physics. Many studies show that both SL and physical understanding and problem-solving skills are essential to enhance students' abilities in STEM learning (Yuliati *et al.*, 2018; Uzpen *et al.*, 2019; and Parno *et al.*, 2020). Moreover, physics is an underlying discipline to understand nature and the world. Historically, understanding celestial motions and the structure of the universe influenced the way people saw the world. Even the view of how nature works are rooted in the aesthetic perspectives of physics: symmetry, simplicity and harmony. Many physicists appreciate the beauty of physics and physics teachers nowadays facilitate students to have a variety of senses such as fun, curiosity and awe from physics. As a means to bring about changes in the future, physics has a great contribution to society. Quantum computing, high-speed transportation, moon exploration and space telescopes are often introduced in the media, and governments around the world are investing huge budgets in the development of advanced science and technology in physics. That is to say, physics is crucial in achieving SL these days.

7.2 EMERGING NEW LITERACIES IN THE SOCIETIES

The American Library Association (ALA) defines digital literacy as "the ability to use information and communication technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills" (ALA, 2022). Today, educators pay much attention to the digital literacy due to the increase in digital tools such as wearable devices (List, 2019). The use of digital tools helps people achieve more convenience in their lives and makes our lives more prosperous through increased access to content. Nowadays, digital transformation induced by current developments in physical sciences and data engineering rapidly changes our lives. A variety of literacies besides SL, such as digital transformation, data, media and ICT, require that we focus on these literacies and their implications for education. With the advent of new digital technologies such as the metaverse and social media, it is becoming increasingly important to obtain and appreciate much information appearing in the digital world. Importantly, students can learn various knowledge and information from a variety of sources and they should be able to know how to create, communicate, and share digital content.

Recently, artificial intelligence has been widely used in many fields and addresses the significance of information management and judgment. The public should be able to understand and appreciate the knowledge acquired from a variety of sources to cope with the vulnerable nature of the digital world, e.g., deep fakes invoked by computer vision and reproduction of biased information due to the internet. Information literacy refers to "the process of capacity building whereby a learner develops the capacity to work independently and socially, and participates in, benefits from and contributes to the information society and the wider global community" (Kong, 2008). Similar to the information literacy, the ability to understand, communicate, analyze and create data as information is called data literacy (McDowall *et al.*, 2021). Mandinach and Gummer (2016) suggested a framework of data literacy for teaching regarding the student's psychological and cognitive development.

Digital tools and media have a great influence on the formation of an individual's image and attitude toward science in addition to acquiring knowledge. Media is an important channel of acquiring information and knowledge and is mainly accessed by digital tools. However, it may lead us to fall into bias or to have negative images of science (Rosenthal, 2020). Media literacy refers to the essential understanding and skills to create, understand and participate in the media in society as a citizen (Thoman and Jolls, 2008; and Buckingham and Rodríguez, 2013). Media literacy can be categorized into four dimensions: media access and use, language and critical comprehension, production and programming processes transforming one's situation through communication (Sánchez *et al.*, 2021).

The aforementioned literacies overlap to some extent and there needs an inclusive term to connect them (as shown in Table 7.1). ICT literacy or MI literacy is often used to synthesize the terms mentioned before. ICT literacy is defined as skills related to the use of technology with skillsets related to the handling of digital information (Lennon *et al.*, 2003). Individuals with ICT literacy appropriately use digital technology and communication tools to access, manage, integrate, and evaluate information; construct new knowledge; and communicate with others in order to participate effectively in society (Scherer and Siddiq, 2019). UNESCO (2013) addresses that Media and Information literacy (MIL) is "to empower people to exercise their universal rights and fundamental freedoms, such as freedom of opinion and expression, as well as to seek, impart and receive information, taking advantage of emerging opportunities in the most effective, inclusive, ethical and efficient manner for the benefit of all individuals." The European Union (2016, 2021) is aware that many citizens lack the ability to exploit the full potential of digital technologies in their everyday lives and focuses on the key competencies

Table 7.1

Various types of literacy [Adaptation from Werts (2008)].

Division	Description
Digital literacy	Digital literacy is more than just the technical ability to operate digital devices properly; it comprises a variety of cognitive skills that are utilized in executing tasks in digital environments, such as surfing the web, deciphering user interfaces, working with databases and chatting in chat rooms.
Information literacy	Information Literacy is defined as the ability to know when there is a need for information, to be able to identify, locate, evaluate, and effectively use that information for the issue or problem at hand.
Data literacy	Data literacy is the ability to read, write and communicate data in context, including an understanding of data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case, application and resulting value.
Media literacy	Media Literacy is the ability to access, analyze, evaluate, & produce communication in a variety of forms. In essence, a media literate person can think critically about what they see, hear and read in books, newspapers, magazines, television, radio, movies, music, advertising, video games, the Internet, and new emerging technology.

to exploit the full potential of digital technologies in their lives, known as DigComp (The Digital Competence Framework for Citizens). OECD also highlights the necessity of ICT literacy centring on science, engineering and mathematics (OECD, 2012, 2018).

The emergence and the stress of other emerging literacies may bring about the debilitation of science education. In appearance, it is likely that SL is subordinate to other literacies or it is unsatisfactory to meet contemporary demands. The aforementioned literacies such as data literacy, information literacy, media literacy and digital literacy seem more relevant to our lives in the present and future. However, such an argument is not adequate since the underlying techniques are rooted in science and technology and a wide range of proficiency and skills relevant to ICT literacy overlap with that of SL (Genlott and Grönlund, 2016; and Senkbeil, 2021).

Moreover, excessive use of digital devices may impede the cultivation of necessary literacy in students. For example, Rosen and Jaruszewicz (2009) stressed the application of appropriate technology aligned with students' developmental levels and Alhumaid (2019) warned us of four ways that may bring about the negative effect of technology-based education as listed deterioration of students' competencies in reading, writing, and arithmetic, which are the three basic skills any student is expected to master; dehumanization of education in many environments and distortion of the relationship between teachers and students; isolation of students in a digital and virtual world that distances them from any form of social interaction; and Deepening of social inequalities between the haves and the have-nots that are students who can possess technology and those who cannot. Even worse, the applied technologies make more profit than the fundamental science. In Korea, the revised curriculum of Korea in 2022 proclaims the important literacy necessary for the future society and allocates more time to digital literacy, whereas it is expected that science classes will be reduced by ten percent or less (Ministry of Education, 2021). It is ironical that science is becoming marginalized by digital technologies, even though the digital environment was brought to us by the development of science and fostering a workforce of science and technology is still important for the future society. It is therefore meaningful to think about the ultimate goal of ICT literacy and SL.

7.3 SCIENTIFIC LITERACY AS A KEY IN THE PRESENT AND THE FUTURE

It seems out-dated to highlight the SL in regard to the rapidly changing societies. Such a viewpoint is from the limited vision of Roberts (2007) identified two different visions of SL: Vision I starts from and focuses on scientific content and scientific processes to learn about corresponding applications later, whilst Vision II focuses on contextualizing scientific knowledge for giving its use in life and society meaning. Conceptual and procedural knowledge of science, including physics, obtained from schools is insufficient to appreciate information in our daily lives and to make appropriate decisions on the problems that we have not encountered before. Digital transformation enables us to obtain a variety of information instantaneously and even access expert knowledge through the internet and social media. Students are not only learners but also teachers who can decide what to learn. For example, if a student is curious about self-driving cars, he/she can search for the keyword self-driving technology and find many video clips explaining the principles of self-driving control, current status of self-driving cars, and even how to make a self-driving unit using microprocessor units. Probably, a student can seek the serious cases warning the risk of self-driving cars such as car accidents, hacking, and ethical and legal issues. The student should be able to access information and knowledge about issues and to appreciate and make a right choice based on what is obtained. Focusing on concepts (Vision I) is not adequate to resolve such issues, but finding out the meaning in the everyday life (Vision II) is more useful to learn and practice science.

As competencies and skills related to the future society are highlighted in many institutions, the concept of SL needs to be revised to reflect these demands. The definition of SL used in an educational setting is based on the references released before the emphasis on digital transformation and the fourth industrial revolution (Miller, 1998; and Laugksch, 2000). Since the new millennium, many research institutions have addressed brand-new skills for the future, e.g., 21st-century skills (Fadel, 2008), future competences (Marope *et al.*, 2019) and core competencies. 21st-century skills provide core subjects and future skills to accomplish three subordinate abilities (life and career skills, learning and innovation skills and information, media and technology skills). It shows a list of 21st-century themes: global awareness, financial/economic/business/entrepreneurial literacy, civic literacy and health literacy.

To cope with the demand for the future, its concepts and principles should be modified considering the possible shape of the future. Project 2061 also mentioned what students should be able to know and do in science and mathematics, such as scientific methods and scientific enquiry (AAAS, 2000). Turiman *et al.* (2012) viewed SL as a foundation for digital age literacy and focused on science process skills such as observation, classification, measurement, inference, interpretation and evaluation of hypotheses. Croce and Firestone (2020) also define SL in line with emerging literacies such as health and media literacy. Choi *et al.* (2011) proposed a new concept of SL to foster the citizens in the twenty-first century, which is composed of five dimensions: content knowledge, habits of mind, science as human endeavor, meta-cognition and self-direction, and character and values. This indicates that SL can correspond to the growing demand for the future by modifying and reconceptualizing the element of SL.

Besides, SL is essential to foster self-directed learning for the future. As a learner-and-teacher, one should be able to access and learn information and knowledge by oneself. Despite the increase in the source of information, many people still rely on the general source for science news and do not tend to doubt the trust of the information according to the report from Pew Research Centre (2017). Only one-in-six U.S. adults are active news consumers who are looking for news and compare them. To do so, students should be able to read science news and understand the meaning of terminologies. In this vein, content knowledge is a key issue to have the capacity for autonomous learning, and school science curriculum should be designed precisely to fit the life and society where students will live.

As knowledge from various disciplines is gradually integrated, SL also needs to be revised to have intersections with various disciplines. For example, solid state physics and statistical mechanics contributed to the development of artificial intelligence in computer engineering and now deep learning is used for calculating the motion and energy of particles at a subatomic level. A deeper understanding of tunneling effect in quantum mechanics led to the development of a scanning tunneling microscope (Binnig and Rohrer, 1987). As a consequence, new physical and chemical devices enable physicists and biologists to manipulate atoms and molecules and finally bring about new discipline (computational physics/biology) and new culture (bio-art and nano-art). Renovations in ICT create new genres of culture, such as animation and plays in the virtual environment and multimedia literature using Twitter and Tiktok.

Moreover, many physics terms are widely used in the media and news. For example, entropy is used to illustrate the diversity of society, to define the status of the financial market, or to explain the value of information on social media (Song *et al.*, 2017; and Garcia-Rubio *et al.*, 2018). Many scientific concepts permeate different levels of society and the overlap between science and other disciplines is growing rapidly (Pearson *et al.*, 2015). Science knowledge is already involved in a variety of aspects of our lives. For instance, high-tech companies have a large proportion of stock and bond market in the world and energy supply is crucial in our lives (black-out will destroy our business and make us trouble). Economy, sociology and many other studies rely on statistical methods developed by science and science is a popular source of art and media such as Interstellar and Gravity.

Looking back at many kinds of literacies, SL is crucial in achieving them. Scientific methods and processes are essential to understand how data is generated and analyzed with respect to data and information literacy. Our concern with physics literacy is embedded within these ideas. For example, data literacy comprises various aspects such as collecting, analyzing, interpreting, implementing, and evaluating data (Suraydi *et al.*, 2020). Such aspects are very much aligned with enquiry in physics. In addition, underlying physical principles in operating equipment, e.g., thermometer and speed meter, are relevant to assessing the quality of data. Appreciation of information and knowledge is closely linked to understanding science terminologies from the media in order to cultivate media literacy. For example, physical terms such as energy, momentum and entropy are often introduced in news articles and social media. To be aware of the benefits and risks of using digital tools, it is necessary to understand how to operate the digital system or device. A broad range of science knowledge and skills are required to accomplish complex literacies. More important is that we should look for ways to connect content physics with every-day life. For example, in the reference to self-driving cars, why not say students can then look for information on momentum, energy, or other topics in physics, perhaps for a project.

7.4 TRENDS IN RESEARCH ON SCIENTIFIC LITERACY

One useful way to have an overview of studies on SL is to conduct a review of literature based on the scientometric approach. The Web of Science, one of the largest academic databases powered by Clarivate Analytics, provides information about the journals and articles officially approved as the internationally prestigious ones. Through the database, I collected a total of 3596 articles dealing with SL since 1992. The detailed bibliometric information of each article was collected and analyzed in different ways: annual production of articles, worldly collaboration, and historiography using R-Studio (Aria and Cuccurullo, 2017). Figure 7.1 represents the annual scientific production of the literature on SL for the last three decades. It is shown that the number of articles dealing with SL is steadily increasing and the annual production was highest at 481 in 2021. About 560 articles are expected to be published by the end of 2022. This indicates the growing interest in SL in the research community.

SL has received much attention spatially as well as temporally. Figure 7.2 represents how science educators collaborate on a study for SL all over the world. In particular, researchers in four countries (U.S., U.K., Australia, and China) are producing large number of publications and are most collaborating with scholars in other countries. With respect to the nationality of authors, English-speaking countries such as the U.S. (3142), Australia (624) and the UK (547) occupy a large proportion of studies on SL according to Fig. 7.2. It is interesting to note that China (722), Israel (144), and Brazil (74) are productive among non-European countries (Fig. 7.3).

The historical direct citation network shows the research trends on SL. For the last two decades, one of the most powerful papers was titled “Scientific literacy for citizenship” by Kolstø (2001). This article highlights the scientific aspects of controversial socio-scientific issues and views SL as a means of resolving these issues. The paper had a direct impact on several articles. Following his article, Hodson (2003) stressed SL to cope with issues due to the globalization, Lee and Roth (2003) proposed new propositions of SL as a community-based practice and Sadler and Zeidler (2004) regarded that socio-scientific issues are fundamental to the notion of SL.

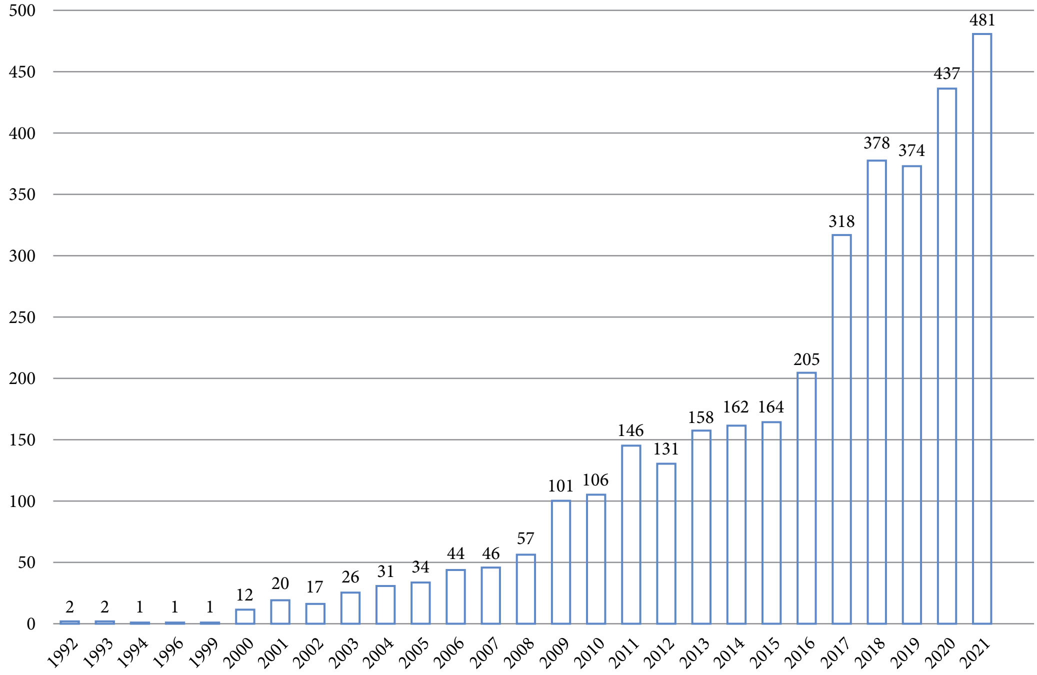


FIG. 7.1
Annual scientific production in relation to the literature on scientific literacy.

Later, the studies on SL are identified with three different topics: socio-scientific issues, the nature of science and argumentation in science. With respect to the socio-scientific issues, Sadler and others became interested in SL in relation to the socio-scientific issues and intended to engage scientific knowledge with students' practice (Zeidler *et al.*, 2005; Sadler *et al.*, 2007; and Sadler and Zeidler, 2009). Regarding the nature of science, Holbrook and others believed that the nature of science is crucial in achieving SL and tried to examine the nature of science as a component of SL through the comparative assessment (Holbrook and Rannikmae, 2007; Bybee *et al.*, 2009; Sadler and Zeidler, 2009; and Bybee and McCrae, 2011). Last, in terms of argumentation in science, Osborne and others stressed SL encompasses logical reasoning which is crucial in the argumentation in science (Osborne *et al.*, 2004; Simon *et al.*, 2006; and Feinstein, 2011).

Figure 7.4 shows the trend of topic by picking up the top four bigrams of each year since the new millennium. The radii of circles mean the number of keywords and the length of lines the period presented in the literature with the least 5 frequencies. For the first 5 years, it is likely that SL studies focused more on basic literacy related to texts as the keywords relevant to reading or writing were

FIG. 7.2
World Collaboration map about research on the scientific literacy.

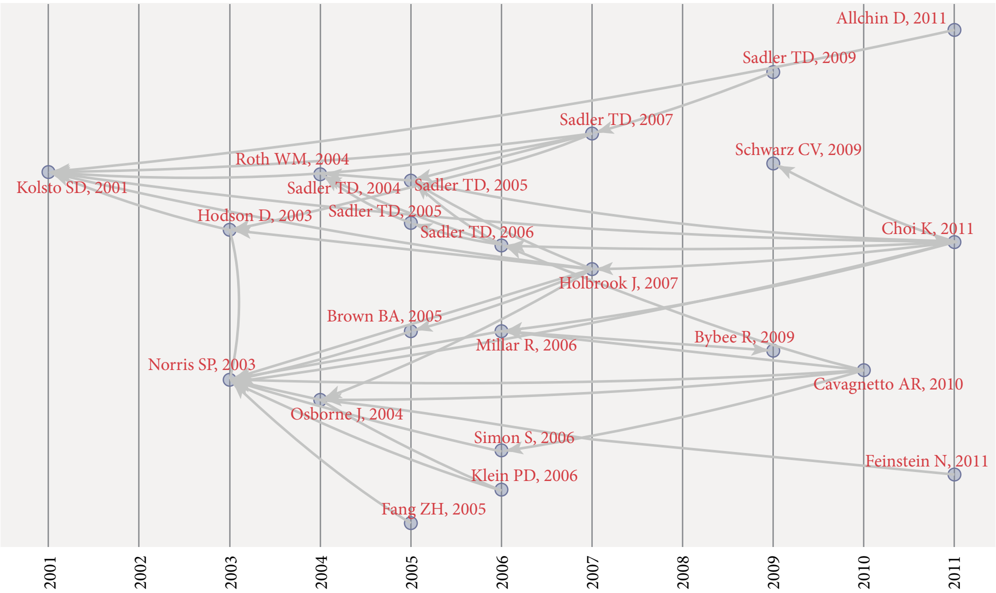


FIG. 7.3
Historical direct citation network for the last two decades.

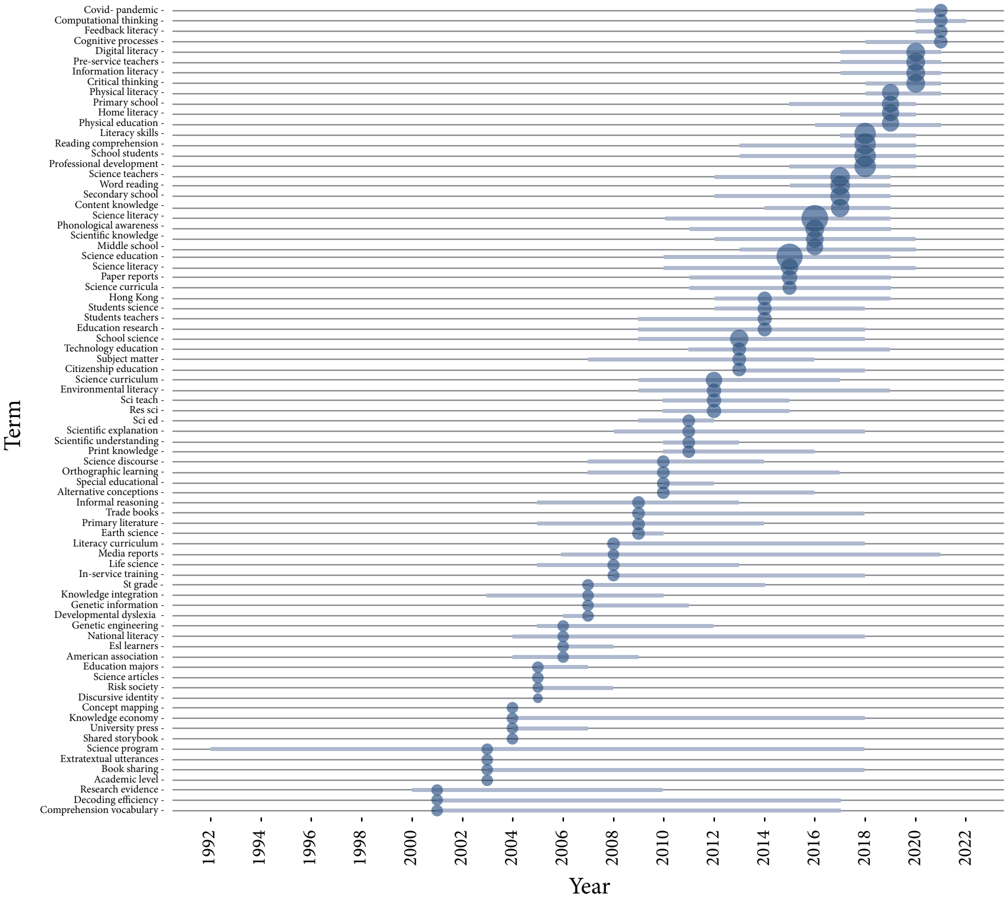


FIG. 7.4
Annual trending topics for the last two decades.

found as listed comprehensive vocabulary, book sharing and science articles. For the next 5 years, biological contents were more addressed, e.g., genetic engineering, genetic information and life science. During the first half of the 2010s, curriculum contents relevant to SL were more heeded, such as scientific understanding, subject matter, science curriculum and school science. For the last 6 years, teacher education for SL has been more heeded in the literature review, e.g., science teacher, pre-service teacher and professional development. It is noted that recent studies on SL concentrate on contemporary skills relevant to the future, such as digital literacy, computational thinking, information literacy and critical thinking. It is likely that such a change reflects contemporary demands on science education. For example, Klucevsek (2017) claimed to bridge SL and information literacy on the ground that there was a large intersection between the two and information literacy-supported SL. Viera and Tenreiro-Viera (2017) stressed that students should be given opportunities to be engaged in learning experiences that promote SL and critical thinking using EdTech. It is crucial to utilize digital literacy as a means of facilitating the cognitive outcomes of students and teachers.

7.5 PHYSICS AS A KEY CONTENT OF KNOWLEDGE IN SCIENTIFIC LITERACY

Physics is crucial in the essence of SL. Foremost, physical theories and laws are fundamental to construct knowledge in the different fields of science. Physics basically illustrates the motions of bodies in mesoscopic and macroscopic systems. As well, physics is involved in any microscopic phenomena: analyzing and predicting structures of new chemicals, interactions of tiny particles among atoms, and electromagnetic influences in organic materials. No one can deny that the emergence of modern physics has brought about a revolution in modern science and technology and civilization in our lives.

Thus, the international comparative assessment entails physics knowledge in multiple ways. For example, PISA articulates the meaning of SL in a sophisticated way and intends to measure students' performances in SL as the content of science (content knowledge), underlying methods and practices used to establish scientific knowledge (procedural knowledge) and understanding of the rationale for the practices of scientific enquiry (epistemic knowledge) (PISA, 2019). With respect to content knowledge, a number of contents related to physics are included as follows:

- Structure of matter (e.g., particle model, bonds)
- Properties of matter (e.g., changes of state, thermal and electrical conductivity)
- Chemical changes of matter (e.g., chemical reactions, energy transfer, acids/bases)
- Motion and forces (e.g., velocity, friction) and action at a distance (e.g., magnetic, gravitational and electrostatic forces)
- Energy and its transformation (e.g., conservation, dissipation, chemical reactions)
- Interactions between energy and matter (e.g., light and radio waves, sound and seismic waves) (Table 7.2).

Procedural knowledge closely linked to enquiry in physics is necessary for achieving SL in PISA. Precise measurement and concepts of variables are often used in physical experiments, and data interpretation and visualization are related to modeling in physics. Epistemic knowledge contains the nature of scientific observation models and theories as well as the value of science. Perspectives on physics are crucial in an appropriate understanding of the nature of science. The nature of contemporary science is engaged with the philosophy of physics. In the twentieth century, a philosophical position of logical positivism was advocated to understand measurement and observation. Observable variables are considered only as physical reality whilst nominal or metaphysical items are to be rejected as meaningless (Williams, 2016). Such an idea was adopted to explain the discrete spectrum of hydrogens

Table 7.2

Aspects of the scientific literacy assessment framework for PISA (2019).

Division	Description
Contexts	Personal, local/national and global issues, both current and historical, demand some understanding of science and technology
Knowledge	An understanding of the major facts, concepts and explanatory theories that form the basis of scientific knowledge. Such knowledge includes knowledge of both the natural world and technological artefacts (content knowledge), knowledge of how such ideas are produced (procedural knowledge), and an understanding of the underlying rationale for these procedures and the justification for their use (epistemic knowledge).
Competencies	The ability to explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically

and the splitting line of spectrum induced by the magnetic field (so-called, Zeeman effect) (Dapprich and Schuster, 2015; Lewis, 2016; and Williams, 2016). The interplay between objects and behavior can be delineated as the theory-laden nature of science and the introduction of quantum hypothesis can be an example of the imaginative and creative nature of science. As well, the basic reasoning such as deduction and induction was also proposed by Aristotle to explain kinematics (Cushing, 1998).

Physics plays an anchor role in achieving literacies pertaining to the future competencies. For example, randomized and standardized distributions of data are based on classical and quantum statistical mechanics. Essential concepts and theories for digital devices rely on physics. Mobile phones, tablet devices and wireless communication are based on electromagnetism. That is to say, key ideas in relation to digital and ICT literacies are necessarily connected with physics.

Regarding the integrated nature of knowledge, physics has given insights to other disciplines. One of the the new perspectives in economics stood upon thermodynamics. For example, the balance price in the free-price market can be explained by the relationship between pressure (price) and volume (quantity of money), and inflation can be defined as entropy formed by gross production, investment and value of materials (Jaber *et al.*, 2006; and Müller, 2007). Key concepts in AI and data science rely on the concepts of statistical mechanics. Hinton (2007), a forefather of AI, advocated a symmetrically connected network with a stochastic decision node, so called Boltzmann machine. According to this algorithm, the total input, z_i, is calculated where the sum of bias b_i, the weights on the connection between i and j, w_{ij}, s_j is 1 if unit j is on and 0 otherwise ($z_i = b_i + \sum s_j w_{ij}$). The probability is given by the logistic function ($P = \frac{1}{1+e^{-z_i}}$). If the units are updated sequentially in any order that does not depend on their total inputs, the network will eventually reach a Boltzmann distribution (also called its equilibrium or stationary distribution). He proposed a new deep learning algorithm using Maxwell-Boltzmann distribution, one of the the key concepts in statistical mechanics.

Even in art, many artists were keen on the theory of relativity and intended to apply physical ideas to their paintings (Parkinson, 2008). Cubists and Futurists were inspired by the dynamic and intertwined nature of space-time (Jho, 2019b). For example, Picasso gazed at a single person or object from various angles, and also combined different appearances into one, like "Weeping woman" and "Guernica." This implies that the depiction of an object is not determined at a fixed point, but different sights are plausible as every system of reference is equivalent with respect to the formulation of the fundamental laws of physics according to the theory of relativity. Balla, one of the founders of Italian futurism, depicted a running dog as if it had many swinging legs. His painting "Dynamism of a dog on a leash" was affected by the chrono-photography, which superimposed different photo frames into one across time. In fact, fast-moving objects leave an afterimage on the retina, making it look like a real dog has twenty legs instead of four. This implies that different motions may leave different phenomena aligned with the principle of relativity and space and time can be visualized in the same coordinate system even though they are distinct and incompatible. Advanced techniques in science and engineering are used for the new genre of art. Nano-art uses synthetic structures with features sized at the nano-meter scale and using quantum optical devices.

Conversely, various disciplines including art may influence physics. Painters came to realize the characteristics of light empirically. For example, Francesco Grimaldi discovered an interesting phenomenon: multiple bright bands were observed when a beam of light went into the dark room through a narrow window. He coined the term diffraction, which originated from the Latin *diffringere*, "to break into pieces." His work was about 150 years ahead of physicists' experiments like Young and Fresnel. Kepler suggested the ordered orbital systems of the solar planets. He used to meet musicians playing instruments on his trip and came to know that there are harmony and rules in music. As such, he postulated that the universe created by God is governed by harmony and order. In this vein, he compared the motions of the planets to music and considered the frequency of the planets to be the pitch of a note. It was explained that Mercury, the planet with the shortest orbital period, would produce the highest notes, and Saturn, the longest planet, would produce the lowest notes (Ferguson, 2013; and Jho, 2014). Even religious dogma plays a role in developing new ideas in physics. In the nineteenth century, Maxwell proposed three primary colors as the basic composition of light, which was inspired by the Trinity, the main doctrine of Christianity (Dharma-Wardana, 2013).

Physics contributed to the progression of education. In the nineteenth century, electromagnetic induction was first discovered by Faraday. He contributed to the development of electromagnetism and electrochemistry by discovering electromagnetic induction, electrolysis and diamagnetism. He delivered a lecture to the public and demonstrated electrical and magnetic experiments to the audience in the theatre (Seeger, 1968). At the beginning of the twentieth century, Tesla made a device producing a high voltage of 200,000 to 300,000 volts and demonstrated the relevant experiment by touching his fingers with the electrically charged objects (Seeger, 1968; Forbes and Mahon, 2014; and Tesla, 2018).

Communication with physicists and the public helps to enhance the awareness of the public about science. Growing interest in science influenced the introduction of science in the public education system. Before the nineteenth century, there was no national system for education. During the eighteenth century, churches provided children with opportunities to learn the basic ability to read, which was called Sunday schools (Power, 1863). The public education system was established over time. In particular, science was included in the school curriculum by the grammar schools act of 1840. In 1870, the Education act was proclaimed to deal with the provision of education in the United Kingdom and explicitly articulated a commitment to national education. On the one hand, the demand on the workforce in science and engineering was rapidly growing due to the industrial revolution. As a consequence, mechanics' institutes were established all over England (DeBoer, 2000). The institutes were later transformed into universities such as Birbeck in the University of London and the University of Manchester. Physicists contributed to the reform of science education. Maxwell and other physicists contributed to the establishment of laboratories and the experiment education in colleges (Mahon, 2004). Later, laboratories and equipment were spread out to schools.

Advances in physics are related to the great changes in music and culture in the modern times (Jackson, 2006). A vacuum valve, a type of diode using a thermionic emission, was first invented by Fleming in 1904. Later, it was influential in the development of transistors and semiconductors. de Forest invented the audio amplifier of the triode vacuum valve in 1912. The triode has three terminals with a control grid that can modulate the flow of electrons from the filament to the plate. It was used for AM radio and the commercial radio broadcasting started in 1920. The advent of new media has made a great contribution to the growth of the music market. Before the radio broadcasting, the phonograph brought about a great change in the music since the disk for the phonograph allowed people to keep and enjoy music at any time. Radio broadcasting spreads music to the nations and the continents (Taylor *et al.*, 2012). Moreover, physics played a significant role in the emergence of new instruments and genres. An electric guitar converts the vibration of strings into electrical signals and can be amplified by loudspeakers. Also, the sound can be altered to perform different timbres or tonal qualities. The mechanically created sound was attractive to musicians and a variety of genres such as rock and roll, heavy metal and electric blue became popular (Charlton, 2014). Even, physics is a good subject for popular culture such as movies and dramas. Science fiction movies tell a story based on time travel and twin paradoxes, and physicists are often introduced as the heroes to save the world or as single-minded obsessive scientists. In addition, the multiverse derived from many-world interpretations is used to depict parallel universes in the movies.

7.6 PHYSICS AS A FUNDAMENTAL BELIEF IN SCIENTIFIC LITERACY

Physics is a cornerstone for the nature of science. The nature of science (NOS) is crucial to achieving SL since it helps to improve understanding of science concepts and to make an informed decisions on personal and social issues as a critical component of SL (NSTA, 2022). Views of NOS play a significant role in acquiring science knowledge, conducting science investigations and making decisions on the relevant issues (Holbrook and Rannikmae, 2007; Michel and Neumann, 2016; and McComas, 2020). Recent studies on the nature of science highlight the epistemic dimensions set aside in the traditional approach to the nature of science (Erduran and Dagher, 2014; Erduran *et al.*, 2019; and Kaya *et al.*, 2019). Aims and values refer to a set of aims in the sense that products of scientific activities are expected to achieve them (Jho, 2019a). Viability, testability, and replicability can be regarded as aims, while simplicity and consistency can be conjoined as values.

In fact, beauty refers to “a combination of qualities, such as shape, color, or form, that pleases the aesthetic senses, especially the sight” (Oxford University Press, 2022). Beauty itself is composed of intrinsic and extrinsic aspects (Sheppard, 1987). Intrinsic properties are related to the formal expressions such as lines, colors, and shapes. On the one hand, extrinsic properties are connected with the subjective judgment or emotion evoked by the forms of an object or an event. The aesthetics of physics encompasses intrinsic and extrinsic properties of beauty, as shown in Table 7.3 (Jho, 2018).

Table 7.3
Intrinsic and extrinsic representation in esthetics of science.

Category	Component	Brief description
Intrinsic	Simplicity	Simple form of theories and explanations
	Symmetry	Invariance or similarities to transformations or rotations
	Harmony	Regularities and order behind nature
	Complexity	Beauty as breaking symmetry and simplicity
	Correspondence	Concurrence between theory and nature
	Unity	Comprehensive nature of theories across context and time
	Invariance	The unchangeable and complete nature of theory
	Coherence	Logical consistency and adequacy
	Visualization	Making the invisible visible
	Abstraction	Expression of invisible form of theory and nature
	Mathematization	Mathematical formulation of theory and nature
Extrinsic	Sublimity	Excellence of beauty with deep affection
	Elegance	Neatness pleasingly ingenious or stylish
	Anew	A new or different affection afar from typicality
	Delight	Pleasure or satisfaction
	Empathy	Understanding and sharing the feelings of others
	Passion	Intense and strong emotion to theory or people
	Inspiration	The state of mentally stimulating
	Mystery	Obscure or puzzling mind to theory and nature
	Splendour	Luxury or grandeur
	Novelty	The quality of being unusual or unique
	Wonder	Surprise or astonishment about theory or nature

With respect to the intrinsic properties, physics intends to represent simple and stable rules or patterns governing the nature: simplicity and symmetry. For example, Dirac (1963) admitted that it was more important to find the beauty in an equation rather than the formula fit for the experiment. A pioneer of quantum mechanics, Werner Heisenberg, assumed that the beauty of natural science reflected the beauty of nature and the ultimate goal of physics is to find out the beauty in physics, not to find out the exact equations (Heisenberg, 1970, 2007). He tried to formulate the uncertainty principle on the ground of the belief about the world governed by the simplicity as a form of beauty as quoted

> "If nature leads us to mathematical forms of great simplicity and beauty—by forms, I am referring to coherent systems of hypotheses, axioms, etc.—to forms that no one has previously encountered, we cannot help thinking that they are "true," that they reveal a genuine feature of nature…" (Thiessen, 2012).

Even the aesthetic preference for simplicity was a ground for the paradigmatic shift in the scientific revolution (McAllister, 1996). Thomas Kuhn, renowned for the structure of scientific revolutions, pointed out that the astronomical revolution was not due to the accuracy of competing theory but to the aesthetic preference about the universe (Kuhn, 2012). In the fifteenth century, geocentric theory was predominant to explain the celestial motions. Copernicus, a Catholic priest, advocated a new system with concentric circles and the sun centered in the midst of them. At that time, the heliocentric system was inferior to the geocentric system in terms of accuracy. However, the geocentric system consisted of a number of epicycles in order to depict the retrograde motion. In terms of accuracy, the geocentric system is more plausible; however, the heliocentric systems are governed by simpler motions. Accordingly, Galileo preferred the simple universe and discovered empirical evidence supporting their viewpoint. That is to say, the aesthetic preference in physics brought about the establishment of new theories explaining the celestial motions (Fig. 7.5).

Symmetry is a prolonged beauty of physics that many physicists have considered before. In Greek philosophy, *symmetria* means balanced, rhythmical and finely detailed human bodies or objects. In physics, symmetry comprises diverse forms of equations or figures. Weyl (1980) and Zee (1999) categorized symmetry into four different aspects: translational symmetry that a particular translation of an object does not change the object itself, bilateral symmetry in which similar anatomical or structural parts are arranged on opposite sides of a median axis so that only one plane can divide the individual into essentially identical halves, rotational symmetry that a shape does not change when an object is rotated on its own axis, and ornamental symmetry that a specific pattern or shape is repetitively found in an object, e.g., hexagons in a honeycomb. Various equations relevant to translational and rotational motion are good examples pursuing the symmetry in physics. Displacement x corresponds to angular displacement theta and velocity v corresponds to the angular velocity omega. Acceleration is defined as the degree of velocity divided by time ($a = \mathrm{d}v/\mathrm{d}t$). In this equation, a and v can be replaced with α and ω respectively ($\alpha = \mathrm{d}\omega/\mathrm{d}t$). In this way, the equations listed in Table 7.4 show the correspondence between translational and rotational motions. Also, Maxwell equations accounting for the electromagnetism look like coupled such as $\nabla \cdot E = \left(\frac{\rho}{\epsilon_0}\right)$ and $\nabla \cdot B = 0$, and as $\nabla \times E = -\left(\frac{\partial B}{\partial t}\right)$ and $\nabla \times B = \mu_0 J + \left(\frac{1}{c^2}\right)\left(\frac{\partial E}{\partial t}\right)$.

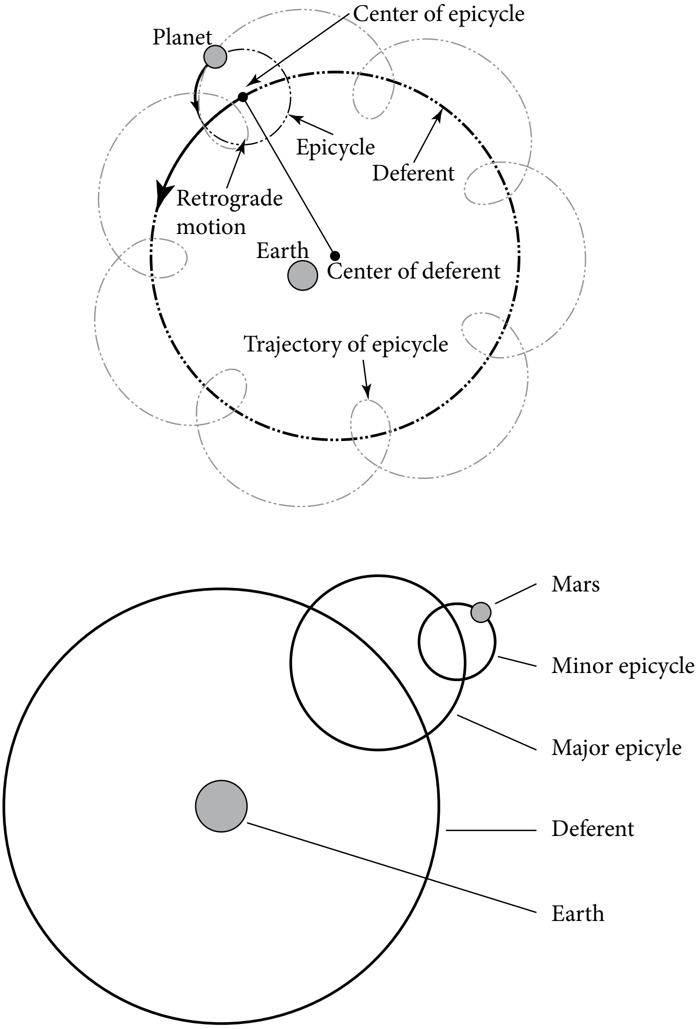


FIG. 7.5
Depiction of retrograde motion using epicycles.

Emotions evoked by natural phenomena, such as surprise or mystery, also play an important role in physics. Physicists use the sublime to indicate a phase transition from solid to gas that bypasses the liquid state and is triggered by an endothermic process just below the critical-point threshold (Crease, 2017). Kant (2007) proposed two kinds of sublimity: the mathematical sublimity that mathematical forms go beyond our imagination or intuition and the dynamical sublimity of experiences overwhelmed

Table 7.4

The symmetric nature of translational and rotational motions.

Translational motion	Rotational motion about a fixed axis
Displacement x	Angular displacement θ
Velocity $v = \frac{dx}{dt}$	Angular velocity $\omega = \frac{d\theta}{dt}$
Acceleration $a = \frac{dv}{dt}$	Angular acceleration $\alpha = \frac{d\omega}{dt}$
Mass m	Inertia I
Force $F = ma$	Torque $\tau = I\alpha$
Linear momentum $p = mv$	Angular momentum $L = I\omega$
Work $W = Fs$	Work $W = \tau\theta$
Kinetic energy $E_k = \left(\frac{1}{2}\right)mv^2$	Kinetic energy $E_k = \left(\frac{1}{2}\right)I\omega^2$

by natural power such as tornado and thunderstorm. In particular, astronomical physicists used to feel sublime or wonder from nature. Besides, delight plays a role in motivating the research of physics. Henry Poincaré witnessed the pursuit of beauty in physics as follows:

> "The scientist does not study nature because it is useful; he studies it because he delights in it, and he delights in it because it is beautiful… intellectual beauty is what makes intelligence sure and strong." (Henry Poincaré, 2010).

For physics, aesthetic nature is crucial in understanding the natural world and theories. Quantum mechanics fundamental to the recent science and technology are also based on the aesthetic standards. A Nobel laureate, Chandrasehkar (1990) addressed scientists seeking theories displaying an adequate conformity in the complexity as a whole. Fermions such as electrons in an atom are anti-symmetric, and any particle is identical. Dominant in force and matters in a microscopic world is super-symmetry (SUSY). SUSY is spacetime symmetry between two basic classes of particles: bosons, which have an integer-valued spin and follow Bose–Einstein statistics, and fermions, which have a half-integer-valued spin and follow Fermi-Dirac statistics (Brading and Castellani, 2003; Mainzer, 2005; and Gangopadhyaya *et al.*, 2018). SUSY is applicable to various fields of physics such as statistical mechanics, quantum field theory, nuclear physics.

Although the symmetric nature of physics is invisible and counter-intuitive, such an idea could be applied to the natural world. For example, the shape of leaf veins and tree branches seems to some extent to be bilaterally symmetric. Symmetry is a basis on the static equilibrium of the natural world and artefacts. The concepts and views of physics on the world are subjects of novels and movies, e.g., rivalry between good and evil and the collapse of different identities similar to pair annihilation

($e^+ + e^- \rightarrow \gamma + \gamma$). Even though the use of science terms in our lives is far from the authentic meaning in physics, a variety of concepts and ideas in physics can be applied in broader contexts.

The aesthetics of physics is essential to NOS and crucial in achieving SL. aesthetic appreciation is a meaningful means to account for a complex nature and is connected with NOS when planning, observing, measuring, collecting, analyzing and inferring (Matthews, 2002). For example, an individual's aesthetic preference and worldview may be adopted to have a look at science as well as art. In addition, aesthetic values can also affect the interpretation of data and decision making since a variety of standards are adopted in decision making to deal with trade-offs and cut-offs (Parmigiani and Inoue, 2009; and Sidarus *et al.*, 2019). Aesthetic appreciation influences not only choosing competing theories but also predicting the patterns from information. In the meantime, scientific reasoning and methods, which are the components of NOS, were adopted, and learning and decision-making are the important outcomes of practices in SL. Both NOS and the aesthetic appreciation influenced understanding science knowledge and making decisions on the issues in which the obtained knowledge and information are engaged.

7.7 CONCLUSION

Advances in science and technology have brought many benefits to our societies. Thus, SL is an important goal of education as it becomes more heeded in many countries. Moreover, the digital transformation induced by development in physical sciences and data engineering rapidly changes our lives and a variety of literacies is being highlighted to achieve the capacity to cope with the future changes as listed digital literacy, information literacy, data literacy, media literacy, and ICT literacy. The surge of alternative literacies has engendered the contraction of SL in the classroom. However, SL is a basis for accomplishing different types of literacies and the meaning and goal of the term has been changed to meet social demands. In this process, physics as bridge various disciplinary knowledge is an underlying cornerstone to achieve the competences for the future.

Even though there are a variety of definitions of SL, it is useful to investigate the international reports and surveys for assessing SL. It is composed of context, knowledge and competencies. The context where we live is more influenced by techniques and devices derived from physics. Even physics has a great contribution to the reform of art, media and culture. The remarkable transitions in mankind were driven by physics. For example, the use of thermal energy was a powerful source of enterprise in the industrial revolution and nowadays, semiconductors and electromagnetism have led to the renovation of industries and societies. Physics is useful to integrate diverse disciplines and to lead a brand-new way of culture. In the twentieth century, idiosyncratic genres of art (bio-art, op-art and nano-art) emerged with the help of physics and new types of culture such as science festivals and concerts have spread out in favor of science. Even, physics contributed to inventing new musical instruments and growing the music market. In terms of knowledge, physics is closely connected with other fields of science and stimulates new ideas in the different societies. For instance, price balance in a free market with P-V graph, and performance evaluation of machine learning using entropy. Even in order to enhance competencies, critical is to cope with the integrating society and interdisciplinary and creative thinking is more important than any other thing people should be able to do. Since physics is to see how the world works and the aesthetics of physics is more fundamental than other components of NOS, understanding of physics is essential to develop expertise for the future. For the last few centuries, physics has contributed to the generation of a creative culture for the public. Even the aesthetic features of physics provide us with opportunities to enhance the abilities of creative problem solving. Creative movements in the films and societies borrow concepts and principles in physics and this is helpful to reconceptualize SL fit for the future changes.

The main point of this chapter is not to uphold the traditional teaching of physics contents. It is true that learning physics is essential to achieve digital or ICT literacies. However, uncontextualized contents may appeal students to engage in their understanding of daily lives. One of the features of a future society is permeability, in other words, crossing borders between science and others. Students should be able to understand the fundamental contents of physics and to bridge physics and other disciplines in an unfamiliar context. They should be taught to apply what they learned to the unexperienced situation and receive more opportunities to solve the problems. In this vein, the value of physics seems more relevant than the authentic meaning of physics contents.

To reflect the growing interests in and concerns about the future society, physics as well as SL should be modified. Meanwhile, a variety of people should take part in negotiating the objectives and common values of SL and different kinds of literacies should be mapped into SL. As only a few concepts are considered in SL for the future, we should extract physical knowledge to be taught to students based on the literature review and experts' interviews. A continuum of physics literacy will contribute to the development of science education in the future schools.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2020R1C1C1005754).

REFERENCES

Alhumaid, K., J. Educ. Soc. Res. **9**, 10–20 (2019).
American Association for the Advancement of Science, *Designs for Science Literacy* (Oxford University Press, 2000).
American Libraries Associations, see https://literacy.ala.org/digital-literacy/ for (2022).
Aria, M. and Cuccurullo, C., J. Inf. **11**, 959–975 (2017).
Binnig, G. and Rohrer, H., Rev. Mod. Phys. **59**, 615–625 (1987).
Brading, K. and Castellani, E., *Symmetries in Physics: Philosophical Reflections* (Cambridge University Press, 2003).
Buckingham, D. and Rodríguez, C., Commun. Rev. Cie. Comun. Educ. **20**(40), 49–58 (2013).
Bybee, R. and McCrae, B., Int. J. Sci. Educ. **33**(1), 7–26 (2011).
Bybee, R. *et al.*, Sci. Educ. **46**(8), 865–883 (2009).

Chandrasekhar, S., *Truth and Beauty: Aesthetics and Motivations in Science* (The University of Chicago Press, 1990).

Charlton, K., *Rock Music Styles: A History* (McGraw Hill, 2014).

Choi, K. *et al.*, J. Res. Sci. Teach. **48**(6), 670–697 (2011).

Crease, R. P., see https://physicsworld.com/a/the-scientific-sublime/#:~:text=Physicists%20use%20sublime%20to%20indicate,below%20the%20critical%2Dpoint%20threshold for "The scientific sublime" (2017).

Croce, K. A. and Firestone, J. B., *Developing Science Literacy in the 21st Century* (Information Age Publishing, 2020).

Cushing, J. T., *Philosophical Concepts in Physics: The Historical Relation Between Philosophy and Scientific Theories* (Cambridge University Press, 1998).

Dapprich, J. P. and Schuster, A., *Philosophy and Logic of Quantum Physics: An Investigation of the Metaphysical and Logical Implications of Quantum Physics* (Peter Lang Edition, 2015).

DeBoer, G. E., J. Res. Sci. Teach. **37**, 582–601 (2000).

Department for Education, see https://www.gov.uk/government/publications/national-curriculum-in-england-science-programmes-of-study for "National curriculum in England: Science programmes of study" (2015).

Dharma-Wardana, C., *A Physicist's View of Matter and Mind* (World Scientific, 2013).

Dirac, P. A. M., Sci. Am. **208**, 45–53 (1963).

Erduran, S. and Dagher, Z. R., *Reconceptualizing the Nature of Science for Science Education*, edited by S. Erduran and Z. R. Dagher (Springer, 2014).

Erduran, S. *et al.*, Sci. Educ. **28**, 311–328 (2019).

European Union, see http://ec.europa.eu/social/publications for "The European digital competence framework for citizens" (2016).

European Union, see https://europa.eu/eurobarometer/surveys/detail/2237/ for "European citizens' knowledge and attitudes towards science and technology" (2021).

Fadel, C., see https://www.oecd.org/site/educeri21st/40756908.pdf for "21st century skills: How can you prepare students for the new global economy" (2008).

Feinstein, N., Sci. Educ. **95**(1), 168–185 (2011).

Ferguson, K., *Tycho and Kepler: The Unlikely Partnership That Forever Changed our Understanding of the Heavens* (Transworld, 2013).

Forbes, N. and Mahon, B., *Faraday, Maxwell, and the Electromagnetic Field: How Two Men Revolutionized Physics* (Prometheus, 2014).

Gangopadhyaya, A. *et al.*, *Supersymmetric Quantum Mechanics* (World Scientific, 2018).

Garcia-Rubio, C. *et al.*, Electronics **7**, 380 (2018).

Genlott, A. A. and Grönlund, Å, Comput. Educ. **99**, 68–80 (2016).

Heisenberg, W., see https://inters.org/heisenberg-beauty-natural-science for "The meaning of beauty in exact natural science" (1970).

Heisenberg, W., *Physics and Philosophy: The Revolution in Modern Science* (Penguin Books, 2007).

Hinton, G. E., see https://www.cs.toronto.edu/~hinton/csc321/readings/boltz321.pdf for "Boltzmann machines" (2007).

Hodson, D., Int. J. Sci. Educ. **25**(6), 645–670 (2003).

Holbrook, J. and Rannikmae, M., Int. J. Sci. Educ. **29**(11), 1347–1362 (2007).

Jaber, M. Y. *et al.*, Appl. Math. Modell. **30**(9), 867–883 (2006).

Jackson, M. W., *Harmonious Triads: Physicists, Musicians, and Instrument Makers in Nineteenth-Century Germany* (The MIT Press, 2006).

Jho, H., J. Korean Assoc. Sci. Educ. **34**(8), 755–765 (2014).

Jho, H., J. Korean Phys. Soc. **73**, 401–413 (2018).

Jho, H., J. Korean Assoc. Sci. Educ. **39**(5), 599–611 (2019a).

Jho, H., Found. Sci. **24**(3), 527–540 (2019b).

Kant, I., *Critique of Judgment* (Cosimo Classics, 2007).

Kaya, E. *et al.*, Int. J. Sci. Educ. **41**(1), 21–47 (2019).

Klucevsek, K., Commun. Inf. Lit. **11**, 354–365 (2017).

Kolstø, S. D., Sci. Educ. **85**(3), 291–310 (2001).

Kong, S. C., Comput. Educ. **51**, 129–141 (2008).

Kuhn, T. S., *The Structure of Scientific Revolutions: 50th Anniversary Edition* (University of Chicago Press, 2012).

Laugksch, R. C., Sci. Educ. **84**(1), 71–94 (2000).

Lee, S. and Roth, W.-M., Sci. Technol. Human Values **28**(3), 403–424 (2003).

Lennon, M. L. *et al.*, *Feasibility Study for the PISA ICT Literacy Sssessment* (E. T. Service, 2003).

Lewis, P. J., *Quantum Ontology: A Guide to the Metaphysics of Quantum Mechanics* (Oxford University Press, 2016).

List, A., Comput. Educ. **138**, 146–158 (2019).

Mahon, B., *The Man Who Changed Everything: The Life of James Clerk Maxwell* (Wiley, 2004).

Mainzer, K., *Symmetry and Complexity: The Spirit and Beauty of Nonlinear Science* (World Scientific, 2005).

Mandinach, E. B. and Gummer, E. S., *Data Literacy for Educators: Making it Count in Teacher Preparation and Practice* (Teachers College Press, 2016).

Marchau, V. A. W. J. *et al.*, *Decision Making Under Deep Uncertainty: From Theory to Practice* (Springer, 2019).

Marope, M. *et al.*, see http://www.ibe.unesco.org/sites/default/files/resources/future_competences_and_the_future_of_curriculum.pdf?fbclid=IwAR1Y3Ac0dNgpHqle02-OwYQ-b-eAZfwEHJ3PwnbUZa8TXIa-PZQpTrMjnhU for "Future competences and the future of curriculum" (2019).

Matthews, P., J. Aesthet. Art Criti. **60**, 37–48 (2002).

McAllister, J. W., *Beauty and Revolution in Science* (Cornell University Press, 1996).

McComas, W., *Nature of Science in Science Instruction* (Springer, 2020).
McDowall, A. *et al.*, Asia-Pac. J. Teach. Educ. **49**(5), 487–502 (2021).
Michel, H. and Neumann, I., Sci. Educ. **25**, 951–975 (2016).
Miller, J. D., Public Understanding Sci. **7**, 203–223 (1998).
Ministry of Education, see https://www.moe.go.kr/sn3hcv/doc.html?fn=09679882cb5c8fa956d3e12c47c94123&rs=/upload/synap/202204/ for "The outline of 2015 revised national curriculum of Korea" (2021).
Müller, I., *A History of Thermodynamics: The Doctrine of Energy and Entropy* (Springer, 2007).
National Science Teaching Association, see https://www.nsta.org/nstas-official-positions/nature-science#:~:text=Nature%20of%20science%20(NOS)%20is,based%20personal%20and%20societal%20issues for "Nature of science" (2022).
OECD, *Literacy, Numeracy and Problem Solving in Technology-Rich Environments: Framework for the OECD Survey of Adult Skills* (OECD Publishing, 2012).
OECD, *Equity in Education: Breaking Down Barriers to Social Mobility* (OECD Publishing, 2018).
OECD, *PISA 2018 Science Framework in PISA 2018 Assessment and Analytical Framework* (OECD Publishing, 2019).
Osborne, J. *et al.*, J. Res. Sci. Teach. **41**(10), 994–1020 (2004).
Oxford University Press, see https://languages.oup.com/google-dictionary-en/ for "Oxford Languages" (2022).
Parkinson, G., *Surrealism, art and Modern Science* (Yale University Press, 2008).
Parmigiani, G. and Inoue, L., *Decision Theory: Principles and Approaches* (Wiley, 2009).
Parno *et al.*, J. Phys.: Conf. Ser. **1491**, 012030 (2020).
Pearson, P. D. *et al.*, Theory Pract. **54**(3), 228–237 (2015).
Pew Research Center, see https://www.pewresearch.org/journalism/wp-content/uploads/sites/8/2017/09/PJ_2017.09.20_Science-and-News_FINAL.pdf for "Science news and information today" (2017).
PISA, see https://www.oecd.org/pisa/pisa-for-development/10-How-PISA-D-measures-science-literacy.pdf for "How does PISA for development measure scientific literacy?" (2017).
PISA, see https://www.oecd.org/education/pisa-2018-assessment-and-analytical-framework-b25efab8-en.htm for "PISA 2018 assessment and analytical framework" (2019).
Poincaré, H., *The Foundations of Science: Science and Hypothesis, the Value of Science, Science and Metho* (Nabu Press, 2010).
Powell, W. A., *Socioscientific Issue-Based Instruction for Scientific Literacy Development* (IGI Global, 2021).
Power, J. C., *The Rise and Progress of Sunday Schools: A Biography of Robert Raikes and William Fox* (Sheldon & Company, 1863).
Ratcliffe, M. and Grace, M., *Science Education for Citizenship* (Open University, 2003).
Roberts, D. A., *Handbook of Research on Science Education*, edited by S. K. Abell and N. G. Lederman (Lawrence Erlbaum, 2007), pp. 729–780.
Rosen, D. B. and Jaruszewicz, C., J. Early Child. Teach. Educ. **30**, 162–171 (2009). doi:
Rosenthal, S., Front. Commun. **5**, 581585 (2020).
Sadler, T. D. *et al.*, Res. Sci. Educ. **37**, 371–391 (2007).
Sadler, T. D. and Zeidler, D. L., Sci. Educ. **88**(1), 4–27 (2004).
Sadler, T. D. and Zeidler, D. L., J. Res. Sci. Teach. **46**(8), 909–921 (2009).
Sánchez, S. L. C. *et al.*, Educ. Rev. **73**(4), 487–502 (2021).
Scherer, R. and Siddiq, F., Comput. Educ. **138**, 13–32 (2019).
Seeger, R. J., Phys. Today **21**(8), 30 (1968).
Senkbeil, M., Educ. Inf. Technol. **27**, 3595–3622 (2021).
Sheppard, A., *Aesthetics: an Introduction to the Philosophy of art* (Oxford University Press, 1987).
Sidarus, N. *et al.*, PLoS Comput. Biol. **15**, e1007326 (2019).
Simon, S. *et al.*, Int. J. Sci. Educ. **28**(2–3), 235–260 (2006).
Song, Q. *et al.*, *IEEE Symposium Series on Computational Intelligence* (IEEE, 2017), pp. 1–5.
Street, B. V., *Literacy and Development: Ethnographic Perspectives* (Routledge, 2001).
Suryadi, Mahardika, I. K., Supeno, and Sudarti, J. Phys.: Conf. Ser. **1839**, 012025 (2020).
Taylor, T. D. *et al.*, *Music, Sound, and Technology in America: A Documentary History of Early Phonograph, Cinema, and Radio* (Duke University Press, 2012).
Tesla, N., *My Inventions: The Autobiography of Nikola Tesla* (Martino Fine Books, 2018).
Thiessen, D., *Bittersweet Destiny: The Stormy Evolution of Human Behavior* (Routledge, 2012).
Thoman, E. and Jolls, T., see http://www.medialit.org/sites/default/files/01a_mlkorientation_rev2_0.pdf for (2008).
Turiman, P. *et al.*, Proc. – Soc. Behav. Sci. **59**, 110–116 (2012).
UNESCO, *Global Media and Information Literacy Assessment Framework: Country Readiness and Competencies* (UNESCO, 2013).
Uzpen, B. *et al.*, Eur. J. Phys. **40**, 035701 (2019).
Viera, R. M. and Tenreiro-Viera, C., Int. J. Sci. Math. Educ. **14**, 659–680 (2017).
Werts, C. E., Educ. Libr. **31**, 6–11 (2008).
Weyl, H., *Symmetry* (Princeton University Press, 1980).
Williams, M., *Key Concepts in the Philosophy of Social Research* (Sage, 2016).
Yuliati, L. *et al.*, J. Phys.: Conf. Ser. **1108**, 012026 (2018).
Zee, A., *Fearful Symmetry: The Search for Beauty in Modern Physics* (Princeton University Press, 1999).
Zeidler, D. L. *et al.*, Sci. Educ. **89**(3), 357–377 (2005).

CHAPTER

8

HISTORY OF PHYSICS AND SOCIO-SCIENTIFIC ISSUES: APPROACHING GENDER AND SOCIAL JUSTICE

Thaís Cyrino de Mello Forato, Isabelle Priscila Carneiro de Lima,
and Gabriela Kaiana Ferreira

de Mello Forato, T. C., de Lima, I. P. C., and Ferreira, G. K., "History of physics and socio-scientific issues: Approaching gender and social justice," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 8-1–8-30.

8.1 INTRODUCTION

Physics is a delightful way of understanding the world, describing and explaining phenomena and rationally and diligently supporting possible choices to face ordinary problems. In a broader perspective, the communities in physics education research, physics teaching, and science education have professed the importance of educating people to figure out contemporary problems, including more complex ones that demand interdisciplinary approaches, which involve not only physics but also different fields of natural sciences, human sciences and social sciences.

Such training, according to Hodson (2010), requires an awareness of authentic sociopolitical action to engage people with appropriate, responsible and effective attitudes in concerns of social, political, economic, environmental and ethical-moral issues. Besides, one can recognize how some interests of individuals, groups, companies, politicians and governments (among others) can contribute to creating social and environmental problems. Hodson (2010) also highlights the need to prepare students to deal with controversial issues in a tolerant and moral way, making decisions about what is right, good and fair, and arguing coherently and persuasively about their conceptions. For this, he defends the development of knowledge, skills and attitudes that support future citizens to investigate different points of view, analyzing and evaluating them, recognizing inconsistencies, contradictions

https://doi.org/10.1063/9780735425514_008

and inadequacies. In this way, the Physics Education Research (PER) community and physics teacher trainers can contemplate such values, defended by several researchers mentioned throughout this chapter, as assumptions for the promotion of democracy and social justice (SJ) and for the training of people, able to promote and share more environmentally sustainable and equitable spaces.

Therefore, it is essential that researchers in physics education engage themselves in discussions about the challenges and problems that societies face, such as the increasing accumulation of greenhouse gases in the Earth's atmosphere, which has led to an acceleration of global warming, intensified by numerous human actions. It has resulted in unprecedented changes in global environmental conditions that are potentially irreversible. Such events have consequences for the health of populations and open up social asymmetries, especially in societies characterized by pluralism (Cortina, 2007).

The research conducted by Tessum *et al.* (2019) exemplifies some of these problems and challenges, such as the environmental racism faced by black and Hispanic people in the United States regarding ethnic-racial disparities in exposure to pollution and the consumption of goods and services. According to the authors, these groups carry a disproportionate burden of air pollution caused mainly by non-Hispanic whites, exposing a disparity between the pollution people cause and the pollution they are exposed to. The authors still denounce the socio-environmental vulnerability of groups more directly affected by the predatory human actions on the planet who have caused the anticipation of extremes in the global hydrological cycle, causing landslides and floods (which aggravate the transmission rates of infectious diseases, especially those transmitted by water and opportunistic vectors).

Likewise, due to possible human action on the environment, the Covid-19 pandemic exacerbated denialist movements such as the anti-vaccine movement and strengthened fascist and Nazi-inspired extreme right-wing movements in different parts of the globe. The pandemic has also opened up social inequalities in several countries in the global south. Many aspects of life in society were affected. The discrepancy in access to treatments, protection, hygiene equipment, as well as the infrastructure for following up with online education became noticeable.

Beyond these challenges, the physics education community is specifically facing conspiracy theory movements in their classrooms, like the flat-earth defenders, the geocentrists, the Christians who deny the Big-Bang Theory, and the climate denialists. Are they prepared or were they prepared to face it? This situation has mobilized researchers, and physics educators to seek transformative changes based on different views by which physics is taught.

This shows that the most vulnerable portion of populations is the most impacted by the drastic consequences of the current model of socioeconomic development and certain predatory practices to the environment. In this context, other urgent topics are brought to science education and the teaching of physics, for example, by

> "[…] the Black Lives Matter movement, which illuminated **racial inequity**; […], discoveries of **Indigenous children buried** at Canadian residential schools, which raised national and global

> **concerns for truth and reconciliation**. Such challenges to humanity may be linked to science and technology, such as about negative health outcomes and **environmental degradation from artisanal coltan mining** in the Congo (Leon-Kabamba *et al.*, 2018) or social media platforms espousing **Eurocentric ideals that infringe on Indigenous identity**, and sovereignty (Matamoros-Fernández, 2017)." (Ibrahim *et al.*, 2022, pp. 33–34, emphasis added)

Colonialist and Eurocentric values, which over the centuries have been configuring a relationship of cultural, ethnic and racial subordination, had affected and affects countless originating and immigrant people on the planet (D'Aambrósio, 1985, 2006; Pingree, 1992; and Gavroglu *et al.*, 2008). According to Aikenhead (2006, 2010), it requires a decolonization of the school science curriculum. Other research has incorporated the concern with the promotion of democracy, regarding the access to education, quality of life, and anti-racist education, for example:

> "Longbottom explores the nature of science teaching if science education is justified in terms of **socio-political goals**. He argues that science education should "contribute to the **advancement of democracy, and so improve the quality of human existence**" (Longbottom, 1999, p. 4). Alberto Rodriguez explores the potential of science education to serve as a **platform for resistance** – a notion only recently beginning to be explored in science education writing, though well established in, for example, **anti-racist education**" (Ahmed *et al.*, 1998). (Reiss, 2003, p. 2, emphasis added).

Inequalities of access, human rights violations, and lack of recognition are also present in the erasure and exclusion of women throughout the history of physics (HP). The report of the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2017) establishes seventeen Sustainable Development Goals (SDG) for Education 2030 Agenda, focusing on the education of girls and women for science (physics included), expressed in "Goal 5: Achieve gender equality and empower all women and girls." Irina Bokova, former Director-General of UNESCO, highlights the enormous disparities and deep inequality in the under-representation of girls in science, technology, engineering, and mathematics (STEM) education, arguing that they are key parts in developing solutions to improve life and to generate "green" and inclusive growth that benefits humanity. Both education and gender equality are fundamental issues for promoting human rights, inclusion, and sustainable development, catalysts to achieve all other SDGs.

Gender inequalities are also highlighted by Reiss (2003), who is concerned about the extent to which physics classes are prepared to train people committed to SJ. Additionally to gender bias, the author points out other important aspects to enable students to critically discuss scientific issues that make them more capable of understanding the scope of physics, but are also aware of its potential for good and bad things. The author emphasizes the different life trajectories, expectations, learning profiles, and individual, class, and economic position differences in certain ethnicities and on science resources to be provided for pupils with physical disabilities:

> "No longer is it implicitly assumed, for instance, that **physics is an activity undertaken predominantly by white middle class males** interested chiefly in car acceleration and the

> motion of cricket balls. More generally, a greater number of teachers realise that the **content of what they teach and the way they teach can turn pupils onto science or off it**. […] Should, for example, **the same science resources be provided for a pupil with a physical disability** (such as severe sight impairment) and a pupil without such a disability? Surely not. But should both pupils receive exactly the same science curriculum? The question is a harder one. And **what of girls and boys? Should they receive identical teaching approaches**? The issue is hotly contested." (Reiss, 2003, pp. 4–6, emphasis added).

These sensitive topics being pointed out throughout this research also bring a not so implicit message: not only who can do physics, but who can learn physics, and who can benefit from the achievements and products of physics.

> "In the UK, for example, differences in educational attainments in science and other subjects are still **strongly related to class and economic position** (Croxford, 1997; Robinson and White, 1997; and Strand, 1999) while certain **pupils from certain ethnic backgrounds continue to underperform** (Gillborn and Gipps, 1996). Whereas **gender inequalities** in the UK are considerably less than in many other countries (Harding and McGregor, 1995), **girls continue to be several times less likely than boys to continue with the physical sciences once they have the option**, while boys are more likely than girls to leave school with no qualifications. […] Social justice is about the **right treatment of others** [what Gewirtz (1998) characterises as the relational dimension of social justice] and the **fair distribution of resources or opportunities** (the distributional dimension)." (Reiss, 2003, pp. 5, 13, emphasis added).

Given this scenario of social injustices, violation of human rights, gender inequality, and environmental degradation (Santos, 2009), it is important to question the social and ethical responsibility of the Physics Education Research (PER) community in identifying the role of physics in the historical constitution of this model of social organization and exploitation of nature and human beings. Accordingly, some questions are posed for us: How have so many inequalities and injustices been naturalized? How is it possible to deal with the contents of mechanics and the history of mechanics without discussing the application of theories and concepts in wars throughout history? How to teach thermodynamics by addressing the history of thermodynamics without dealing with the effects of industrialization on the environment and the exploitation of human labor? How to address any episode in the history of physics (HP) without discussing its role in the construction of human cultures and the socio-political organization of society? How did physics development contribute to many advances and achievements for several people, but also brought exclusion, hunger, and misery for others? How to prepare physics teachers to deal with such challenges and understand how physics concepts are linked to the sensitive themes of SJ?

The context briefly described above favored the strengthening of SSI in the 2000s, as a possibility of adding socio-political actions involving such themes to STS and STSE approaches (Zeidler *et al.*, 2002, 2005; and Zeidler and Nichols, 2009), as presented in Sec. II of this chapter. Resonating with these ideas, the research community in the history, philosophy, and sociology of science in science education was also pursuing similar goals for the teaching of physics and science, in a significant and expressive movement to approach epistemic and non-epistemic aspects of the nature of science (NOS) (Matthews, 1992, 2014; Metz *et al.*, 2007; Rudge and Howe, 2009; Allchin, 2011; Heering and Höttecke, 2014; Bagdonas and Silva, 2015; and Forato, 2018).

Accordingly, this research explores the potential of socio-scientific issues (SSI) to address these sensitive and urgent topics involving the promotion of SJ when addressed within the framework of HP. The SSI provides a theoretical-methodological foundation for the intellection and debates on content brought by historical narratives (Metz *et al.*, 2007) that might be constructed revealing the educational potential of such an interface. This means that in view of the perspectives defended here, this chapter intends to propose a theoretical framework by combining SSI and HP in the physics teaching.

As an example to apply this theoretical framework, a short narrative on Lise Meitner's contributions to the nuclear fission theory and some gender issues involved in the historical episode will be addressed. This is because it clearly presents injustices related to tackling gender bias, which is increasingly urgent within the Physics Education Research (PER) community. Despite this episode may be well-known for many historians of physics, from the elements mobilized by the SSI discussions, it is possible to identify how the scientists' decisions—their values, religion, the context of a war, and the gender issues involved in the episode—implied undue liability, violation of human rights, and severe consequences for society.

Thus, this theoretical framework applied in the episode seeks to offer subsidies to prepare contextualized approaches, in Freire's perspective, aiming at a transformative education, which can change people and people can change the world (Freire, 1970, 1998; and Santos, 2009). Together with a historiographical perspective (D'ambrósio, 2021) and inspiration from the ethics of cordial reason (Cortina, 2007), it may also provide a basis for elaborating historical narratives capable of mobilizing SJ debates. Such an approach also seeks to offer contributions to pre-service physics students, which may influence their future practices.

8.2 SOCIO SCIENTIFIC ISSUES IN SCIENCE EDUCATION

An international bibliographic review was carried out on the use of the Socio Scientific Issues (SSI) applied to Physics teaching, based on History of Physics (HP) approaches, covering 1980–2020. The objective of this review was to identify theoretical and practical approaches that propose interlocutions between SSI and the HP in PER. So far, however, the review has revealed a certain gap in the current literature between these two themes. An expressive number of dissertations, papers, and books on SSI focusing on the educational areas of Chemistry, Biology, and Science were found. Therefore, only a few focus on Physics teaching and not necessarily using historical approaches.

The discussions, in general, are focused on health, quality of life, climate change, that is, considering research in biology, chemistry, and sciences, and interfaces with other fields. Among the few studies with SSI and physics interfaces, some adopt historical episodes without offering a theoretical-methodological discussion or/and sometimes without mentioning SSI.

In Shapin and Schaffer (1985), a classic work on the historical controversy between Thomas Hobbes and Robert Boyle over Boyle's air-pump experiments in the 1660s, the authors analyze conflicts over the value and propriety of experimental methods. The approach explains a dichotomy between the social and the epistemology of science in search of ways and solutions to produce knowledge in natural philosophy. In the 17th-century English context, such reflections also involved the search for solving political and religious problems. Although the work does not bring a theorization about what the community currently considers as theoretical and methodological bases on SSI, nor the now urgent themes sensitive to SJ, the proposal incorporates emblematic aspects of SSI by bringing contextualized debates about the relationship between epistemic practice in the sciences and social, political, moral, and ethical aspects of the context.

Ann-Marie Mårtensson-Pendrill (2006) presents a sensitive didactic proposal on the relationship between physics, physicists, and politics, in the context of the Manhattan Project, implemented in a teacher-training course. Using various materials, such as the autobiography of many physicists who worked in that period, the proposal offers an opportunity to meet the physicists as persons dealing with difficult complex ethical problems. The author used examples of physicists struggling with the ethical questions of the role of physics and physicists and sought to promote a richer understanding of the context of atomic and nuclear physics and physicists and their relationship to society, including more interesting and thought-provoking discussions on the responsibilities of scientists. Even if she has not entered theoretical or methodological foundations, she classifies her rich experience as an SSI approach, which she considers multidisciplinary in character.

Bagdonas and Silva (2015) have also used the HP to discuss the interface between science and religion, beyond the naïve views perpetuated in the school environment, which foster stereotypes. The authors argued that during the Cold War period, the cosmological controversy between the Big Bang and Steady State Theory was tied to political and religious arguments. They present a didactic sequence developed for and applied in a pre-service physics teacher-training course on the history of science. After studying the historical case, pre-service physics teachers discussed how to deal with possible conflicts between scientific views and students' personal worldviews related to religion. Although the proposal may raise controversial issues without the need to explicitly defend certain positions or disapprove students' cultural traditions, Bagdonas and Silva do not propose the SSI approach.

In the same way, Forato (2018) has presented a didactic proposal to the contextualized teaching of Newton's Law of Universal Gravitation by discussing Isaac Newton harmonizing science, God, and alchemy in a neo-Platonic world view, connected with his intellectual environment in Cambridge, aiming at exemplifying how personal values of a thinker also impact science. The author argues that by providing the experience of these contents in which the students' religious or atheist options are respected, teachers can acquire elements to respect the beliefs and positions of their future students, always pursuing an inclusive attitude towards different points of view. According to her, reflections about such contents can expand epistemological discussions in classrooms and promote respect for diversity to approach human rights. However, the author does not mention the contributions of this historical episode to discuss SSI.

Like Mårtensson-Pendrill (2006), other studies adopt HP controversies for discussing SSI and do not present theoretical and methodological justifications, without constituting a gap or a problem. Even with an understanding of the contributions of such options, this research intends to offer a systematic understanding of how SSI and HP could contribute to physics teaching, and a proposal of how combining both can provide potential pedagogical benefits to Physics classrooms, to deal with one of the sensitive themes involving SJ already presented. In addition to the benefits that historical controversies offer to discuss NOS (Oliveira *et al.*, 2018), it is proposed that the epistemic characteristics of the construction of SSI for the school environment allow for advancement in relation to these and other sensitive and urgent topics linked to the promotion of SJ.

The SSI approach emerges from discussions of the Science, Technology, and Society (STS) and Science, Technology, Society, and Environment (STSE), adding to these movements the demand for discussions on controversial topics covering social problems and necessarily requiring explanations about the values and ethical aspects involved in judgments (Zeidler *et al.*, 2002, 2005; and Zeidler and Nichols, 2009). For Zeidler *et al.* 2005 (p. 371), “the SSI approach represents a reconstruction and evolution of the STS model that provides a means to not only address societal implications of science and technology but also to tap into students’ personal philosophies and belief systems.”

The SSI approach overcomes several criticisms directed at STS, such as the fact that traditional STS proposals or approaches do not explain the ethical dimensions of science, students’ moral and emotional development, nor do they explore the pedagogical power of well-founded discourse and argumentation, or even the NOS and the epistemological debate of science. The advancement of SSI is precisely at this point: to encompass the entire discussion inherent to the STS theoretical framework and additionally consider the ethical, moral, and emotional dimensions of subjects and objects (science) (Zeidler *et al.*, 2002, 2005; Zeidler and Nichols, 2009; and Zeidler, 2014), providing a connection between the various axes that contribute to the development of scientific knowledge.

For Zeidler and collaborators, the

> “SSI therefore does not simply serve as a context for learning science but rather as a pedagogical strategy with **clearly defined goals**. Certainly, knowledge and understanding of the interconnections among science, technology, society, and the environment are major components of developing scientific literacy; however, these interconnections **do not exist independently of students’ personal beliefs**. It is our stance that STS(E) approaches can be remodeled and substantially improved **by adding an essential missing component-consideration of each student’s own moral and ethical development.**” (Zeidler *et al.*, 2005, p. 360, emphasis added).

The SSI approach, therefore, allows addressing controversial issues and problems with a higher level/degree of complexity. Not only from a scientific and technological point of view (which often divide opinions in society), which contributes to establishing divergent opinions but "similarly" reasonable within certain logics and cultures, but also from significant or influential social groups, who raise explanations and conflicting solutions to the same issue or problem, from significant or influential social groups (Levinson, 2006). These controversies based on scientific principles and concepts do not present simple or unique conclusions and may involve risks; they are subject to public discussions and influenced by economic, political, moral, and ethical issues (Zeidler *et al.*, 2005).

Besides, socio-scientific controversies offer critical and reflective discussions for the teaching of physics that is intended to be emancipatory (Freire, 1998; and Santos, 2009) as they are linked to numerous social factors resulting from the application of scientific and technological principles and practices (KolstØ, 2001; and Sadler and Fowler, 2006).

The essential characteristics to recognize a controversy are systematized by Levinson (2006), p. 1204:

> 1. when people start from different premises, hold different key beliefs, understandings, values, or offer conflicting explanations or solutions that are rationally derived from the premises (Crick, 1998; Oulton *et al.*, 2004; and Wales and Clarke, 2005); 2. when it involves a substantial number of people or different groups (Bailey, 1975; Inner London Education Authority, 1986; and Crick, 1998); and 3. when the issue is not capable of being settled by appeal to evidence (Stenhouse, 1970; Stradling, 1984; and Wellington, 1986).

In this context, the SSI, considered demanding discussions on controversial topics covering social problems and necessarily requiring explaining the values and ethical aspects involved in judgments, provides a series of formative objectives to be explored and valued in Scientific Education. It is argued that educational proposals addressing SSI can promote SJ, democratic access to scientific concepts, and the overcoming of passive and noncritical postures in the face of dilemmas.

Besides, SSI can mobilize skills, values, and attitudes characteristic of different cultural, philosophical, moral, and religious traditions (Hodson, 2020; Sadler and Zeidler, 2004; and Macalalag et al., 2020)) and can help students in cognitive, social, political, moral, and ethical development (Millar, 1997; Hammerich, 2000; KolstØ, 2001; and Sadler, 2004). Moreover, they support explicit and reflective discussions about NOS (Rudge and Howe, 2009) in its epistemic and social aspects.

Zeidler *et al.* (2005) consider that to foster the formative objectives pointed out above, it is essential to develop pedagogical proposals involving the SSI that allow reflection on their theoretical structures, enhancing the development of moral, emotional, and argumentative aspects inherent to a dilemma and that involves the structure of scientific knowledge. To this end, the theoretical framework for the SSI approach in the classroom is proposed to promote functional Scientific Literacy (SL) (Zeidler *et al.*, 2005; Zeidler and Nichols, 2009; and Zeidler, 2014). For the authors:

> "In this conceptualization, **functional SL**, in contrast to more traditional notions of SL that are more technocratic in nature, **is dynamically mediated by personal cognitive and moral developmental considerations**. These considerations include factoring in character and cognitive and moral development and include the use of (but may not be limited to) cultural, discourse, case-based, and nature of science issues." (Zeidler and Nichols, 2009, p. 50, emphasis added)

In this theoretical framework, elements such as NOS issues, classroom discourse issues, cultural issues, and case-based issues proposed in an SSI approach promote functional SL when interacting with students' personal cognitive and moral developments. Thus, this SSI approach also aims to overcome the criticisms pointed out for the STS mentioned above.

Chowdhury *et al.* (2020) justify the importance of SSI in science education to promote citizenry and synthesize the expected characteristics of students for the promotion of the desired citizenry; they are personally responsible, participatory, justice-oriented, and politically concerned. Based on the literature, the authors identify attributes associated with SSI, contributions, and barriers to the promotion of citizenry in the context of science education. Among which one can cite: socially inserted scientific contexts oriented to local, national, and global issues; perception of complexity in various values, ethics, and morals; promotion of student participation through a trans-curricular "poorly structured" context, in the sense that SSI proposes an open and complex issue without a practical, fast, unique solution.

The need for an epistemological theoretical matrix for addressing SSI is also pointed out by Levinson (2006), who aims to develop a conceptual basis for the controversial SSI teaching model for high school students based on categories of "reasonable disagreement," the "communicative virtues," and "modes of thought. In this structure" Levinson (2006) provides a typology of levels of disagreement (reasonable disagreement), based on epistemological considerations, that reflect whether people are attached to the same values or different values; differences in priorities about the same values and/or different interpretations about a problem.

For the author, it is reasonable to disagree based on rational justifications, although only rationality may not consider aspects of social, humanistic, and ethical justice. The communicative virtues are necessary dispositions on the part of the subjects to discuss a controversial subject, considering the role of narrative reporting in their sustaining, to engage students in reasonable disagreement and allow dialogue through difference. Finally, Levinson (2006) explains that the modes of thought are based on the thoughts and experiences of the subjects that can better illuminate the disagreements, and may be of the "narrative" or the "logic-scientific" type. The "narrative" in which they involve the voices of the participants plays a vital role in transmitting meaning in a scientific disagreement and contributes to closing the gap that may exist between personal needs and contexts and so-called "scientific" solutions in coping with problems. The "logic-scientific" type is based on scientific evidence and followed by the perspective of science:

> “The logic-scientific mode deals in general causes and their establishment and tests for empirical truth on one hand, and the narrative mode, constructing stories on the other, are interwoven in seeking to convince […] in the context of seeking to give validity to a point of view.” (Levison, 2006, p. 1215)

These two modes of thought are not invariably incompatible, but both are ways of structuring experience to explicate reasonable disagreement.

Saunders and Rennie (2013) defend an SSI theoretical framework focused on developing ethical thinking, contributing to the recognition of other worldviews and multiple identities, including cultural, ethnic, religious, and gender perspectives that can be explored and considered in the resolving of SSI in our science classrooms’ (Saunders and Rennie, 2013, p. 261). Based on a literature review, they structure elements that make up a pedagogical model that aims to structure teachers’ thinking in exploring a controversial SSI. To compose the theoretical framework, the authors identify four traditional elements—consequences, harms and benefits, rights and duties, virtue—care based, right to choose—and add a new so-called pluralism:

> “This notion is further developed in this paper where it is argued that explicit consideration of pluralist aspects can provide a richer view on ethical perspectives. We suggest that raising awareness of other worldviews and identities should not be ignored or marginalized in the resolving of SSI in our science classrooms.” (Saunders and Rennie, 2013, p. 257)

According to the authors, the model has several contributions to teachers and students. The application of the model for ethical inquiry in terms of its use as a pedagogical tool, which assisted teachers in their practice and confidence in addressing SSI in their classrooms, helped the teachers to develop a more substantial pedagogical base to support their teaching and learning about SSI. It was in a way that engaged and motivated students and in doing so moved towards developing their own and their students’ scientific literacy, such as both the teachers and students’ knowledge base about ethical frameworks and ethical decision-making.

Like Saunders and Rennie (2013), Yap (2014) argued that ethical frameworks could be an effective means to explore SSI. The author proposes using ethical frameworks that incorporate Christian values to enable students to face controversial dilemmas in socio-scientific issues. As ethical structures, he considered rights and duties (deontological), maximizing benefits (utilitarian), making decisions for yourself, virtues, and Christian ethics/values, in which he assessed the effectiveness of using ethical references as a pedagogical strategy to facilitate students’ critical thinking, informal reasoning, argumentation, and decision-making skills. Yap (2014) justifies that he used Christian virtues and values to develop the proposal in a Christian school and concludes that the use of ethical frameworks in socio-scientific education as a teaching and learning tool reinstates the importance of incorporating values in science education and establishes a tangible link between moral considerations and scientific literacy.

Both the contributions of Saunders and Rennie (2013) and Yap (2014), when proposing to construct pedagogical models that explore ethical thinking in depth, are relevant, as they emphasize the importance of this thinking from various references guiding the teacher on how to proceed in addressing an SSI in the classroom. However, it is important to assess the context in which a particular religion is elected as a reference for the approach. The proposal may be appropriate for the proposed context. Still, it is necessary to assess to what extent this could conflict with groups of other cultures and ethnicities for whom values of a specific religion were imposed. To what extent could this mean violating the human rights of these people?

Considering the plurality of religious beliefs and values present in societies, it is important to assess which perspective to base topics involving religion could promote respect for all faiths, including the understanding that empathy and respect for differences is the basis for promoting human rights and SJ. Faced with globalization, the migratory flows of people around the planet, and the plurality present in various social contexts, all these people live with big cultural differences in many cases. For example, another sensitive point already mentioned in the introduction is the gender issue. The role of women in society is substantially different in some cultures. How could SSI be an approach to providing support to address such sensitive topics?

8.3 SSI AND HP TO PROMOTE SOCIAL JUSTICE

There is an extensive body of literature on the last decades arguing about the numerous benefits that HP can offer to the teaching of physics, in teacher training, and in elementary school. For example: to humanize the sciences, revealing its personal, ethical, cultural, and political aspects of the community; to favor critical thinking; to give meaning to laws, concepts, formulas, and equations; to improve the scientific and cultural training of the teacher; to understand the structure of the sciences and the space they occupy in the intellectual system; to expand the democratic control of scientific activity; to know the plurality of scientific methodologies accepted in each epoch and each area of science; to understand scientific knowledge as a result of a human construction, inserted in a historical and social process; to understand epistemic and non-epistemic aspects of the NOS, the role that experiments play in the generation and establishment of scientific knowledge; to understand the historical and interdisciplinary character of the development of physics in interface with other areas of knowledge, among others. (e.g., Matthews, 1992, 2014; Metz *et al.*, 2007; Rudge and Howe, 2009; Allchin, 2011; Heering and Höttecke, 2014; Bagdonas and Silva, 2015; and Forato, 2018).

Furthermore, as discussed in the previous section, the SSI approach also allows mobilizing aspects that promote human rights, such as anti-racist education (Reiss, 2003), and prepares students for responsible and well-informed social interaction (Zeidler *et al.*, 2005).

This research argues that in addition to those benefits pointed out in Sec. II, the interface between HP and SSI allows addressing sensitive themes involved in promoting SJ, in the sense of approaching unfair situations in the classroom, on teacher training and basic schooling. Accordingly, the PER community can develop it to support and inspire actions, materials, didactic proposals, and curriculum reviews. In addition, studying HP in this direction favors the discussion of its influence on the processes that contributed to a world organization with so many situations of human rights violations. Therefore, it presents some contributions of Ubiratan D´Ambrósio, Paulo Freire, and Adela Cortina, which resonate with these purposes.

Several episodes show the erasure of women, collaborators, helpers, and other peripheral peoples in the construction of the sciences and reveal possible reasons such narratives can contribute to maintaining relations of subordination, for example, the imposition of Eurocentric ideals among several colonized peoples. This allows, for example, a reflection on how biased historical narratives can be used by projects of cultural domination (D'Aambrósio, 2021).

According to Oliveira *et al.* (2018), some studies adopt HP narratives that highlight the use of historical controversies, exploring their potential to discuss ethical, social, political, economic, and personal aspects intrinsic to scientific practice in proposals for approaching SSI. These works propose the debate of controversial themes among groups of students about historical controversies as a strategy to engage them, mainly for learning aspects of NOS. However, even though they address SSI, little is said about the sensitive themes from the point of view of developing a structure of ethical thinking, human rights, and SJ in the terms proposed in this chapter.

A robust and interesting historiographical perspective to guide the sensitive themes of SJ is the Ethnomathematics program that was established as a field in the 1970s by Ubiratan D'Ambrosio (1985, 2006), considering the development of mathematics and natural sciences, including physics, concerning the cultural, political, social context, and economic forces that shape the world. Inspired by the ideas of Paulo Freire (1970) in Brazil and other thinkers in "peripheral" countries, D'Ambrosio understands HS as the history of the human species in search of survival and transcendence in the various environments it occupies.

In search of education that can reconcile development and sustainability[1], D'Ambrosio points to recognizing the relationship between knowledge systems and human values and advocates the role of researchers and educators to think together, ethical values, and transdisciplinary knowledge. From this perspective, D'Ambrósio advocates an Education for Peace, able to lead the human being to reach the state of real consciousness, only possible when knowledge and human behavior are in solidarity. The author recalls the role of mathematics and natural sciences in the development of technologies, the basis of the current way of life, in which there are immense social injustices, political systems of subordination, and discrimination, also fostered by inequalities in access. In this sense, HP must understand the evolution of human knowledge, permeated by the arts, sciences, values, religions, and behaviors, obviously relating and influencing each other.

Under the current civilizing scenario, D'Ambrósio (2006, 2021) proposes historiographical and methodological alternatives that lead to a history not imbued with a Eurocentric determinism

---

[1] In resonance with some complex problems approached in the introduction.

favoring the maintenance of this *status quo*. He recognizes that the scientific production of peripheral countries is different from that of central ones, being subordinate to them, given that inequalities in infrastructure create a barrier to the effective work of peripheral countries with border issues. Advocating that education is a non-neutral activity and considering its role in the construction of a more just and egalitarian society, he proposes the search for new directions with the greater purpose of guaranteeing the survival of the planet and civilization. Thus, he argues that it is up to the historians of sciences to recover knowledge, values, and attitudes of the originary and colonized people, often relegated to a lower plane, ignored and sometimes even repressed and eliminated, which may be decisive in the search for these new directions. As an example of recognition and integration of the ways of knowing of original inhabitants who have suffered colonization, Aikenhead (2006, 2010) advocates the respect for the Aboriginal culture in Canadá, such as their beliefs, costumes, spirituality, and Indigenous ways of knowing nature. The author integrated a project to decolonize school science and mapped out a rationale to integrate Indigenous *sapientia* into the school (Eurocentric or Western) science curriculum in the Province of Saskatchewan's curriculum renewal. "Decolonizing school science **begins** at the stage of "acceptance" and succeeds at the stage of 'integration.'" (Aikenhead, 2006, p. 393, original emphasis)

The historiographical perspective outlined by the Ethnomathematics program aims to promote human rights and SJ by uncovering the historical roots of the sensitive themes proposed by this research. Only a historiography that allows us to understand the role of physics and science in the constitution of this predatory model of society can bring reflection and awareness and lay the foundations for educational actions in the formation of new generations.

One can start by asking what is the role of science and technology in all this? What is the role of physics in establishing rationality that historically grounded these unequal relations? For Adela Cortina (2007), the Enlightenment rationality of the 19th century is no longer sufficient for the intellection and confrontation of real problems, which are transdisciplinary in themselves, and are constituted because of human and nature exploitation, in the models of development of industrialization and injustices in the distribution of resources and products of the sciences.

Adela Cortina (2007), from the indignation at the suffering of others and the concern with the reality of significant social and economic asymmetries, proposes the investigation of the sense of justice, responsibility, and care for the other. Defending fair coexistence in societies characterized by pluralism, she analyzes the scope and possibility that this feeling of responsibility, care, and cordiality can be learned. When proposing the foundations of civic ethics, in which she seeks to establish minimums of justice with implications in the different areas of public life, Cortina defends the principles of the experiential ethics of care, for example, human concern in the face of its context, not only seeking "what we should do," but "why we should do it":

> "[…] helping oneself and others to empower and improve their lives, through care, thus contributing to the human capacity to fulfill life as something valuable; life as a reality worth living." (Cortina, 2007, pp. 223–226)

Cortina proposes that it is impossible to know justice only by "pure" rationality but for the reason that considers esteem, admiration, and compassion.

The bases proposed by Cortina are resonant with those of D'Ambrósio for HS in education, especially when thinking about themes sensitive to human rights violations, as commented in the introduction of this chapter. The feeling of powerlessness in the face of so much injustice reinforces the need to fight for a genuinely democratic and emancipatory science education based on an ethic of cordial reason (D'Ambrósio, 1985, 2006, 2021; Freire, 1970, 1998; Cortina, 2007; and Santos, 2009), which allows one to look at the other, mobilizing feelings of care and empathy.

Thus, it is proposed that the choice of the historical episodes and the construction of their narratives, to mobilize the SSI in the teaching of physics, make it possible to bring these aspects involved in the ethics of a cordial reason. It pursues the promotion of SJ for vulnerable populations and groups, fostering a transformative and liberating education capable of evoking emotions and feelings of empathy. Among exclusions, unfair systems, vulnerabilities, and prejudice are the gender issues demanding to be approached in a fair education system.

To mobilize such feelings, the Meitner ́s episode intends to promote discussions within the scope of the right to access scientific spaces, and the influence of gender identity for this. Besides gender issues, values and beliefs are revealed by religious prejudices guided by political interests, the choices scientists and politicians have made in the war context, and how much the values of the state and institutions are decisive for a naturalized selection of the profile of scientists who stand out.

This proposal of approximation between HP and SSI does not forget the concern with PER, to bring physics closer to girls and women who want to follow this career. It is known that many actions have been planned and executed in this direction. In the international literature, it is possible to know about initiatives in schools and universities to reduce the difference between the number of girls and women to boys and men, especially in physics courses and in the development of research in the area. Using historical narratives (Metz *et al.*, 2007) can be an effective strategy to present the long-forgotten contributions, victims of this widespread erasure, of women scientists in all areas of knowledge.

There is an urgent need for a diverse, plural science project, different from what HP based on Eurocentric determinism generally promotes. According to Schiebinger (1999, p. 37), the modern science project excluded hundreds of women. That exclusion is about the prohibition of women in science courses at the end of the nineteenth century, when classical physics was already established. Also, and above all, it is about consequences of this for the current model of science. It is not only about having the space to develop research. It is about offering science new views and the possibility of further questions based on a different perspective of facing social and scientific problems. Feminist criticism of science does not only refer to the fact that there are few women. It refers to the model of science reproduced in the laboratories, the structure guiding the elaboration of research protocols, which is based only on the reality of this group representing this colonial science, centered on a uniform group with specific demands.

The highlight of feminist criticism of the still prevailing science model is the possibility of thinking of a plural science. A science in which all groups' demands are considered and presented in the discussions and in the elaboration of this science. According to Arrazola (2002, p. 71–72), scientific knowledge suffers from a "sexist deviation." "A science situated, says Ilona Löwy (2000), opens the way to another conception of objectivity, of universality, which includes diversity, criticism, passion, contestation, solidarity, and responsibility."

If science is considered as an enterprise developed by different identities and narrated from the perspective of non-hegemonic groups, there will be new readings, problems, and interpretations to construct scientific knowledge. The idea is not to make science more subjective or more "feminine" but to build a science with new perspectives and, consequently, more comprehensive, which Keller (2006), p. 15 calls new objectivity. As Harding (1996), p. 15 mentions, "Feminists have stakes in a successor science project that offers a more adequate, richer, better account of a world, in order to live in it well."

Therefore, the assortment of concerns already explored in this chapter resonates with the ideas pointed out by Paulo Freire (1970, 1998) for liberating and transforming education (Santos, 2009). Also, the historiography proposed by Ubiratan D'Ambrósio (1985, 2006, 2021) considers the cultural, political, social context, and economic forces that shape the world. Moreover, with the sense of justice intended by Adela Cortina (2007), proposing a science education based on cordial reason, responsibility and care for the other. All these characteristics were also emphasized by the SSI literature, presented in the section above.

Accordingly, it is argued that addressing SSI using HP allows

- discussing and understanding epistemic and non-epistemic aspects of the NOS, considering the values of students;
- understanding that physics is influenced by the socio-historical context, while it influences it;
- understanding social and personal aspects of scientific practices;
- approaching adequate episodes of HP allows an anti-racist education, aiming at promoting equity, and historical retraction and reparation, as well as approaching Indigenous Originary people *sapientia* on their Cosmogonies;
- understanding how the historical development of physics and its interface with other areas of knowledge contributed to the constitution of global and local exclusionary and unfair systems, and so many situations of human rights violation;
- developing empathy based on a cordial reason, which considers SJ, humanistic, and ethical aspects, and generates commitment and engagement;
- understanding how the HP influenced the clash between forces that shape the unfair world;
- approaching gender issues, supported by feminist criticism of science, can promote the understanding that it is not only about who can do physics but also about the advances and improvements offered by a more plural group—which includes diversity, criticism, passion, contestation, solidarity, and responsibility—doing science;

- giving voice to the students of a plural ethnic-racial group, considering his ethical, aesthetic, ecological, moral, educational, cultural, and religious values, to experience a transformative education;
- reflecting on the responsibility of those who make physics, who research physics education, who teach physics, and who make scientific dissemination of physics. What is the relationship between the view that has been fostered about physics, its history, and SJ?

It is defended that by discussing HP episodes within the framework of SSI, this new knowledge is being built together, and by the students, considering all different values they bring, it is possible to promote a transformative education. Practices and research committed to SJ and human rights require giving voice to the subjects by welcoming the multiplicity of perspectives. Therefore, it is important to understand the individual and PER community responsibilities and train people engaged in social transformation, that requires a transformative education and implies strengthening principles of freedom, the democratization of access, equity, historical retraction and reparation, and scientific literacy that is truly liberating (Freire, 1970).

In this sense, the study on the academic trajectory of Austrian physicist Lise Meitner (1878–1968), from the perspective of feminist criticism of science, offers sufficient elements to discuss how to explore SSI in HP. Besides these discussions already presented and discussed in Lima (2019), this is an episode located at a sensitive and remarkable political moment: the rise of Nazism in Europe and its effects on the production of scientific knowledge. It is important to remember the neo-nazi movements that are spreading in the world today and causing severe polarization in some countries.

The example of Lise Meitner, although located at a point in time and in a specific space, can motivate reflections to be moved to many other spaces and times when other women tried to insert themselves in research institutions. Who decides who can and cannot do science? How long will it be necessary to discuss and recall all possible arguments to justify that the gender of researchers does not make them less capable?

8.4 BUILDING AN EXAMPLE: LISE MEITNER AND SSI ON PHYSICS TEACHING

Lise Meitner was born in Vienna in the late 19th century into a Jewish family. She grew up in that city, attended schools, and at the age of 14 ended the school period for Austrian women at that time (Sime, 1996). The influence of Lise Meitner's gender identity throughout her academic career is a fruitful field to connect science, technology and society, and therefore, to SSI. To teach physics with this episode, it is possible to encourage postures that value democratic access to scientific knowledge, especially as scientists, foster discussions on social justice, and consider the beliefs of students and teachers in the debate on the subject. With this, it is possible to encourage students to adopt a position in the face of social problems that involve gender identity[2] and, in similar cases, racial and class issues, etc.

It is possible that a portion of the scientific community does not associate the outcomes of Lise Meitner's trajectory as being influenced by the scientist's gender identity. This is another element that should be considered and that strengths SSI. The denial of the scientific community regarding the harms and influences of the patriarchal system that surrounds the scientific enterprise is a strong point to be added to the analyses regarding the group directing and making decisions within science, from the first conceptions about scientific knowledge to the present day.

This attitude of denial of the prejudice evident within science must reach Physics Teaching critically so that students can know flaws in the science model hitherto presented to them. A model that values the presence of a specific group of scientists, a non-diverse group, and that definitely removes women, young people, black people, LBTQIA + from scientific careers. In addition, for this reason, the academic trajectory of Lise Meitner serves to address discussions on the values of scientists and ethics in research. The historical episode itself represents an opportunity to understand the role reserved for women in educational and scientific institutions in Germany in the early twentieth century and to study the country's historical and political context during the rise of the Nazi government.

Her trajectory leads us to reflect on the role reserved for academic women at any time of the HS. Without universalizing the term woman, here she is a Jewish woman who could access the spaces of education and science in Berlin at the beginning of the twentieth century. The daughter of a Jewish father and Jewish mother, Lise Meitner was born in Vienna in 1878. Meitner had the support of her relatives to study, starting and finishing her first studies in schools in Vienna. In 1878, Austrian women still organized themselves in the struggle for rights such as voting, as did other groups of women throughout Europe and other parts of the world. Among these rights not yet conquered was that of not being able to continue studying until admission to the university. Women should study up to a stage of basic education and learn only what is necessary, what is useful for them to become housewives, serve their husbands and be exemplary mothers Sime (1996).

Meitner attended school regularly until she was allowed to. Women could not advance and join the Gymnasium[3], as the Austro-Hungarian Empire understood it as an unnecessary expense. For the Vienna government, girls did not need to learn algebra, for example, because it was not necessary to develop so-called feminine skills in the future. The reasoning required in mathematics classes should be taught only to boys (Sime, 1996).

---

[2] To avoid anachronistic analysis, it is sufficient to guide that discussion with the logic of the binary gender structure. It is enough to understand the culture of people who identify as male and female to analyze the expected behaviors and the places occupied by the society of that time. Although we know that the concept of binary gender is already outdated, to analyze Lise Meitner's academic trajectory, the differentiation between feminine and masculine contemplates and provides subsidies for our arguments in the discussion.

[3] Gymnasium was a level like high school (to use current terms), the level that prepares students for college.

One of the justifications used for this was linked to the difference between the bodies of adolescents, based on biological determinism, to interfere with the access to education of young women. According to Clarke (1884), girls should interrupt their studies at their first menstruation, when the body begins to develop characteristics related to sexual development. According to the German Women's Educational Regulation, the girl's physiological specificity determined her time at school. Those who were middle or upper class could have tutors at home, while those of the less favored class ended up dedicating themselves to the field. Here it is evident that the rulers intended to use the women's own bodies as a mechanism to prevent them from continuing their academic careers.

Thus, at this moment, they received what was called Entlassungs-Zeugnis, something like a certificate of dismissal from the school. Lise received her certificate on July 15, 1892, and from there she was to return home, wait for her fiancé, prepare for her marriage and dedicate herself to the life of a future mother (Sime, 1996).

Even without a formal mechanism for this, society still uses the body of (cisgender) women as a regulation to prevent any career advancement in several countries. The grounded cis-hetero concept of compulsory motherhood made it impossible for cisgender women to pursue any career, including scientific ones, as mothers and wives or housewives. There is an explicit relationship between possible scientific careers and the place reserved for cisgender men who will not experience "events" associated with their biology. This discourages women from maintaining a work routine in academic and professional spaces, a sensitive issue to reflect on SJ and the bodies of cisgender women scientists. Thus, it is argued that there is no liberating and transformative education without detachment of the work capacity from the biological condition.

A science built on the patriarchal bases and that guides the constitution of its characters in the sexual division of labor does not present itself as fair and does not promote gender equity. Additionally, if other identity markers intersect here, such as race or sexuality, we will see that exclusion is even greater.

Since 1867, the Universities in Austria have been open to men of any class, origin, or religion. Some women tried to approach the university but were received as an unofficial audience, somewhat informal. They were not well received and did not receive titles. However, women resisted, and groups led by school principals fought for girls' access to preparatory courses and universities. In 1891, the girls were able to attend a Gymnasium for girls, which was called Madchengymnasium, but no guarantees that they could take the exam that would take them to universities, the Matura (Rife, 2017).

In 1897, the government allowed women access to the faculties of philosophical sciences, such as letters and sciences, of Austrian universities and, a few years later, to enter medical schools. Now, with guaranteed access, the government should repair the damage it caused, as to the gap in the training of these girls. In addition, justice required that teachers had university degrees, but how could they acquire this title so quickly if they still had to stay a few years in school to complete the Gymnasium? At that time, the Empire allowed women to take Matura without necessarily attending classes at the Gymnasium (Sime, 1996).

This is the first aspect of Lise Meitner's academic trajectory that presents us with an SSI. The difference in access to education for boys and girls. State values at that time, that country's concept of SJ, and decisions based exclusively on gender differences are characterized as issues that can raise complex debates and divide opinions in the classrooms. Based on their set of principles, some believe that historically, this should be the most coherent measure, and some can identify the serious problem of inequality of access.

The role of HP in physics classes is also to mobilize students to analyze critically the path that leads science characters to their achievements. Understanding the impediment of advancing in school implies delays and, often, dropouts in pursuing careers such as the scientific one. Investigating the root of the problem of the absence of women among researchers offers an understanding of why women are still so few nowadays. From this aspect, the SSI that brings Lise's trajectory explains how much the culture of the society in which she was born and grew up influences decision-making within sectors of the same society and consequently can interfere in the scenario of future scientists.

From this first aspect of the analysis of this biography, we bring two elements highlighted in standpoint feminism, defined by Haraway (1988): the look from the subjugated groups and the non-neutrality of science. It is only when the perspective of analysis of biographies like this is changed that we can see aspects such as impediments as indispensable factors in writing narratives. In the Meitner case, it is necessary to look at the facts from her standpoint to understand the absence of women in the sciences. To discuss the low presence of women in physics, for example, is to realize that there is a historical delay in the admission of these women in schools first and then in universities and work environments. This analysis is guaranteed when we understand the role of feminism from a perspective or standpoint, which allows us to look at these underrepresented groups.

Consequently, with the impediment to continuing to study, Meitner entered the University exactly 5 years later, if there was no prohibition. In this perspective, we identify the historical problem of the absence of women in the field of science and technology. Meitner's late entry into university is a key element for us to understand the little presence of the female gender in S&T productions.

If we compare Meitner's reality to that of some important physics characters who had a very similar background to her, we can confirm how this delay is inherent in women's careers. For example, Erwin Schrodinger (1887–1961) and Wolfgang Pauli (1900–1958) entered universities in Vienna and Munich, respectively, at the age of 18 and were not prevented from pursuing their studies at the time of entering higher education. Meitner entered the university in 1901 at the age of 23 and obtained her doctoral degree in February 1906 (Sime, 1996), becoming the second woman to obtain a doctoral degree in Vienna. Even with the permission granted, few women chose to enter universities, let alone go further and complete a doctorate.

Formal prohibition was, until 1892, a reality in Vienna. Nevertheless, informal prohibitions were and still are a reality for underrepresented groups in these spaces. Currently, no law prohibits girls from choosing courses in physics, mathematics, engineering, etc., but it is clear that the number is still minimal. This fact is explained by Mafia (2002) and called an informal obstacle, referring to the lack of need for laws to become aware of spaces reserved for women in certain places. However, although there is no formal impediment, there is a naturalization of the idea that women should not occupy these spaces. Women themselves are sometimes convinced and internalize a discourse of incapacity.

In an interview with Thomas Kuhn in 1963 (Meitner, 1963), Lise Meitner was asked whether she had faced any problems as a woman in an environment where almost everyone was a man (with sporadic exceptions for female students). In her response, Meitner said that many of her colleagues did not want her to work in chemistry and were not allowed access to some places. In an article published in *Physics Today*, Meitner spoke again about the situation: "I went to Berlin for further studies and presented myself to Planck to attend his lectures. He was very friendly, but clearly astonished; he said, 'You have a doctor's degree, what more do you want?'" (Meitner, 1960, p. 20)

Additionally to being prevented from moving forward for a long time, the few women who managed to access the spaces occupied mainly by men faced the prohibition of access to collectively used places in universities. This directly implies the impossibility of expanding contact networks and new possibilities of work. Here, once again, the episode helps to understand the format of science intended by scientists. Science is a social enterprise that aggregates values and knowledge from a specific group: cis, heterosexual, European men, and from a well-defined social extract. There are specific and well-reserved places with pre-defined conditions to date for those who do not follow this description. Even today, mainly men occupy the experimental areas. Laboratory work is aimed much more at men, as well as the highest positions, the most important places (Lima, 2019).

Only in 1908 were women legally accepted into the universities of Berlin (Sime, 1996). Thus, the situation was as follows: women were legally participating in university activities as students. Activity in the laboratories was allowed, but Meitner had no salary and no job title. In 1912, Meitner and Hahn received an invitation to work at the Kaiser Wilhelm Institute (KWI) in Berlin Dahlem. Hahn was offered a position of scientific associate and the title of Professor, being responsible for a radioactivity section of the first laboratory in Germany and a great annual salary. Meitner was very welcome as a guest and without remuneration (Sime, 1996, p. 45).

Meitner worked hard, had written several articles since 1907 to date and was beginning to present herself as a researcher in theoretical physics, independent of Hahn. She had already published dozens of papers in important journals, yet the Institute leaders did not offer her a position. As already reported, the presence of women was allowed, but in a non-disguised way, the control and impossibility of giving spaces to them were always present. For 4 years, she worked as a volunteer associate without a position or remuneration. Even with so much dedication and competence to develop works as brilliant as those of her colleagues, what she received immediately after a year was an invitation from Max Planck to be his assistant. Finally, Meitner had her first paid job, correcting his students' exams.

The following year, in 1913, Meitner became an associate at the Institute. At the time, she would earn a salary, but much lower than Hahn's. The radioactivity session was at the Hahn-Meitner Laboratory, and Meitner's salary was three times lower than Otto Hahn's (Sime, 1996). After becoming an associate of the institute only in 1919, seven years after its opening and twelve years after Meitner arrived in Berlin, she was granted the title of KWI professor. However, in 1922, she was granted her Habilitation and gained her *venia legendi* (right to teach).

Still, in 1914, Hahn and some of his colleagues from the Institute were called to the war. Meitner took over the laboratory alone and continued conducting the research. Despite being concerned about Hahn and his exposure in the field, Meitner managed to finish some works and published them, all with Hahn's name as the first author. It is unknown whether this was a requirement of the laboratory or Hahn himself. Still, it is evident that even assuming the leadership of the laboratory, Meitner could not appear to the scientific community as the first author of her own works.

It deprives women of the power of speech, exclusively due to their gender identity. According to Keller (2006), this causes a "female self" to become an "androgen node," a "non-man" and this is overvalued. In this case, it is possible to identify one of the mechanisms that can perhaps justify the erasure of contributions of so many women within the spaces of scientific knowledge production. It is known that other women also worked, researched and developed their work in these institutions. However, few of these productions survived attempts at erasure.

The academic trajectory of Lise Meitner from the perspective of feminist criticism of the known narrative highlights this series of aspects related to gender identity that can be discussed in terms of SSI. From these elements, it is possible to connect to Chowdhury *et al.* (2020) discussion about the importance of SSI in scientific education and how it promotes citizenry. By knowing the debate about gender and the influences on women's careers (as well as race issues are also determinants in the careers of black people), students can develop their argumentation skills and, with that, become personally responsible, participatory, justice-oriented, and politically concerned. Moreover, issues related to the political moment experienced in Germany in 1930, its implications for the development of science, both concerning what it was proposed to study (research of interest to political groups), and who was authorized or not to conduct these studies, can be explored.

In the third decade of the 1900s, the Nazi party, which in 1933 had Adolf Hitler (1889–1945) as chancellor, gained the support of many people from Germany, who lived in a depressing post-war period, a devastating crisis period. As an alternative to the country's progress, the National Socialist German Workers' Party was created, of which Adolf Hitler was its leader.

Although Meitner had become a Protestant in 1908, she was of a Jewish family, and for this reason in 1933 she was suspended from the University. Her situation was no more serious than that because she was an Austrian citizen. This still protected her, and although a little safe, she could no longer legally contribute to the institute she coordinated. Between 1933 and 1938, Meitner continued to work, but always in secret, avoiding appearing publicly in those spaces.

Despite the worsening political situation in Berlin, it was in 1933 that Hahn and Meitner began a new project. It is possible to say that there was a beginning to the studies of Lise Meitner and Otto Hahn that culminated in Nuclear Fission. This moment was the Solvay Conference of 1933 and its discussions on the directions of research on the structure and properties of the atomic nucleus. One of the projects was entitled "New transformation processes when uranium is irradiated with neutrons,"[4] published in *Naturwiss* in 1936. This is a good indication that before Lise Meitner had to flee Germany, the works she conducted and published, together with Hahn, were already heading for what, two years later, would be the first publication on the interpretation of the phenomenon of Nuclear Fission.

In March 1938, Hitler announced the Anschluss, annexing Austria to the territory of Germany. Therefore, the Nazi government would persecute any Austrian person of Jewish origin. A Nazi chemist from the KWI, Kurt Hess, denounced Meitner's presence at the Institute, which after being communicated to Hahn, reached Meitner. The denunciation also reached the Reich Research Council. On the night of July 12, Meitner wrote to her friend Elisabeth Schiemann: "I had exactly one and a half hours to pack, to leave Germany after thirty-one years" (Scheich, 1997, p. 161).

Meitner, when fleeing Germany, left behind her workplace, her career, and a place won with the confrontation and, above all, the understanding of her ability to collaborate equally with men. It took a while for that space to be minimally comfortable for her, but it did not take long for her to be forced to leave. It is the extremist policy that is decisive for the direction of science and that will reserve oblivion to the work conducted by Meitner and so many other characters targeted by the Nazi Policy of Extermination (Lima, 2019).

This scenario of conflict and authoritarian governments in Europe also helps to understand how political movements are decisive for understanding the changes in the objects of research and the strategies established by governments to strengthen their assets from sectors such as the scientific. Furthermore, to understand how scientists' identity and political adherence, in addition to their lack of commitment to ethics, also interfere in the direction of this science. People who agreed with Nazi practice remained in their research groups, performing their activities with minor disruption.

In the end, who were the winning scientists? Moreover, what is left for the losers? In the midst of all this, it is necessary to recognize the potential of this episode regarding its discussion in terms of SSI. The social and political context in which Lise Meitner's academic trajectory is inserted allows us to promote a discussion in physics classes about the ethical frameworks in socio-scientific education. In addition, it allows incorporating values into scientific education and establishing a tangible link between moral considerations and scientific literacy. Along the same lines, SSI can allow students to develop in the social, political, moral, and ethical spheres (Millar, 1997; Hammerich, 2000; KolstØ, 2001; and Sadler, 2004).

Meitner left Berlin and took refuge in Stockholm, in a laboratory where she was not in any condition to continue conducting the work started at her institute. She did not have material, did not have

---

[4] In the original: *Neue umwandlungsprozesse bei Bestrahlung des Urans mit Neutronen.*

the equipment and could not even get into the research groups that worked there (Sime, 1996). Between July and January 1939, she communicated with Hahn through letters, and they talked a lot about the work they were doing until the moment of her escape. They were arguing about bombing uranium atoms using neutrons. Meitner offered her contributions in writing, and Hahn conducted the experiments in Berlin. However, on January 6, Hahn and his new co-worker, Strassman, published in *Die Naturwissenchaften* a five-page article, some graphics, and an experimental chemist's explanation of the detection and behavior of uranium irradiation using neutrons producing alkaline earth metals. Meitner's name did not appear in the article.

About Meitner's feelings in early 1939, in a letter to her brother, she wrote

> "Unfortunately I did everything wrong. And now I have no self-confidence, and when I once thought I did things well, now I don't trust myself. The Swedes are so superficial; I don't fit here at all, and although I try not to show it, my inner insecurity is painful and prevents me from thinking calmly. Hahn has just published absolutely wonderful things based on our work together (…) And much as these results make me happy for Hahn, both personally and scientifically, many people here must think I contributed absolutely nothing to it - and now I am so discouraged. although I believe I used to do good work, now I have lost my self-confidence." (Sime, 1996, p. 255)

In a way, she had no self-confidence, no pay, and was betrayed by her work group. Meanwhile, scientists and research groups worldwide reproduced the experiments proposed by Hahn and those that would lead to the results published by her and Frisch. Meitner was already 61 years old and had no prospect of future work. These are more elements that help us think about how the political moment favored Meitner not to have her name printed in the first article on the division of uranium. These are also aspects that allow us to think about personal motivations.

Then, in the same year, Meitner and her physicist nephew Otto Frisch described, on two pages, the physical explanation for what was being observed in Otto Hahn's laboratories in Berlin: the rupture of the surface tension of the atomic nucleus and the generation of kinetic energy due to the loss of mass. Then, the article explains the fission process:

> "On account of their close packing and strong energy exchange, the particles in a heavy nucleus would be expected to move in a collective way which has some resemblance to the movement of a liquid drop. If the movement is made sufficiently violent by adding energy, such a drop may divide itself into two smaller drops (…) It seems therefore possible that the uranium nucleus has only small stability of form, and may, after neutron capture, divide itself into two nuclei of roughly equal size. (…) Therefore, it seems possible that the uranium nucleus (…) can, after capturing the neutron, divide into two nuclei of approximately equal sizes. (…) These two nuclei will repel each other and should gain a total kinetic energy of c. 200 MeV., as calculated from nuclear radius and charge. (…) The whole "fission" process can thus be described in an essentially classical way" (Meitner and Frisch, 1939, p. 239).

Since Meitner and Hahn's papers were published in 1939, their names have begun to be nominated for the Nobel Prize because of studies on nuclear fission. Between 1939 and 1945, considering the Chemistry and Physics Prizes, Meitner was nominated 9 times and Hahn was nominated 18 times. The nominations were from important names in science, such as Arthur Compton, James Franck, Dirk Coster, and Niels Bohr. In 1939, the year in which the publications were being discussed, and the experiments reproduced around the world, the Swedish chemist Theodor Svedberg nominated the names of Hahn and Meitner to share the Prize. The reasons for the nomination were "It seems that sharing the Prize between Hahn and Meitner for the discovery of uranium fission or in common for their work with uranium fission products should not be questioned. Therefore, the sharing of the Prize could also be proposed to a great extent for the totality of their common work in the field of radioactivity." (Rife, 2017, I. 8368)

Theodor Svedberg, who in 1939 had strongly recommended that Meitner and Hahn share the prize in 1941, wrote to the Nobel Committee for Chemistry as chairman saying that Hahn had done essential work for the "discovery" of fission, while the work of Meitner and Frisch had not been extraordinary. Understanding that the work conducted by Meitner was not an extraordinary one is to disregard the importance of the work published in early 1939, and moreover to erase the three decades of work together with Hahn in Chemistry and Nuclear Physics. There is no explicit justification for this, but it is undeniable that there is an attempt to erase Meitner's contributions from all sides (Crawford *et al.*, 2008; and Lima, 2019).

In 1945, Otto Hahn received the Nobel Prize for 1944, "for his discovery of the fission of heavy nuclei." It is worth highlighting and discussing how Hahn refers to the entire process of studies and carrying out works, and the weight he attributes (or not) to Lise Meitner's contribution to the realization of his experiment. When reading the article published by Hahn in Scientific American in 1958, entitled "The Discovery of Fission, ' the oblivion and absence of Lise's name as part of the process and an even more present collaborator than Strassmann is notorious." When reading the article, one can find excerpts like the following:

> Not being physicists, we thought of uranium's atomic weight (238) rather than the number of its protons (92). Subtracting the atomic weight of barium (137) from that of uranium, we guessed 10 l as the atomic weight of the other fragment.(...). Immediately after our paper appeared, Meitner and Otto R. Frisch came out independently with their historic publication showing how Niels Bohr's model of the atom could explain the cleavage of a heavy nucleus into two nuclei of medium size. Meitner and Frisch named the process "fission." (Hahn, 1958, pp. 82, 84)

At no time does Hahn refer to the collaboration of his colleague during her work at KWI, nor does he mention the letters exchanged when Meitner was in Stockholm. He does not mention his questions about the strangeness of the phenomenon and his search for Meitner's opinion as a theoretical physicist. It is also curious how he refers to the work of Meitner and Frisch as having "came out independently," without presenting any connection, once again, of Meitner with his manuscripts (shown to her in half, by the way).

When recognizing that Meitner was also responsible for studies on the atom and the physical interpretation of the phenomenon of nuclear fission, the scientific community and the European press did not give her due credit. Hahn's name came out on the front line; he was credited with the honor of being a Nobel Prize winner, for example.

However, when the application of fission theory exploded, at the first opportunity to blame someone for the harm of the bomb, the person who acts on the front line is no longer the German man who did science in his laboratories. "Who prints the newspapers is the female scientist, Jewish, Lise Meitner" (Lima, 2019, p. 157). Using Sedeño's definition of the principle of female co-participation (Sedeño *et al.*, 1999, p. 211), this responsibility attributed to Meitner is interpreted as another male characteristic in the persecution of women who stand out in these areas. "If a woman does something wrong, it is typical of her sex, of all women, but if one does it well, it is just an exception." In the atomic bomb case, Meitner did not even participate in the project, but she was held accountable because, in this understanding, she is more susceptible to error.

Even though she no longer had any identity with the Jewish community, as she had long since converted to Protestantism, Meitner was remembered as the Jewish mother of the Atomic Bomb. To what extent the political justification is related to the fact that she is not Aryan is enough for this case is an issue that deserves to be discussed. Otto Frisch also wrote the work, was also the author of other articles on the nuclear fission process, and belonged to a Jewish family; however, he was not accused or exposed to the cover of the newspapers at the time.

The political situation around the world was dramatic. The explosion of the two atomic bombs in Hiroshima and Nagasaki by the US increased the number of those responsible for the conflicts. At that moment, science took the place of a protagonist. War resources gained a new and powerful element counted in the amount of energy for destruction. Lise Meitner, until then forgotten, was on radio programs, being interviewed by the former American First Lady Eleanor Roosevelt, and listening to comments that blamed her from people like the president of the United States at the time, Harry Truman (Rife, 2017).

In his tenure, Harry Truman was the president who experienced the beginning of the conflicts in the former Soviet Union, which triggered the Cold War. Currently, as a consequence of these clashes, new geopolitical conflicts, now between Russia and Ukraine, take place in the media and divide the opinions of civilians (not to mention the parallel wars normalized by the Western media, and very little is known about people being massacred in different places on the planet). Once again, the world is experiencing an explicit and politically articulated moment of grand proportions, but with disastrous results, which undoubtedly benefits from advances in technologies and scientific studies with warlike purposes.

Assessing the context in which the theory of nuclear fission was established without considering the relations between SSI is relegated to oblivion to the fundamental role of scientists in projects in favor of armament. However, it is unfair to forget that some of these scientists, like Meitner, had not even agreed to collaborate on these projects. Even more serious, associating her name with the post-nuclear bomb war scenario is, at the very least, a disservice to HP records. The political, social, and economic context arising from these two moments throughout history (1945 and 2022) offers another element for SSI to enter physics classes, not forgetting the role of science as a culture, as a product of personal relations, and as an instrument of negotiations, successes, and failures in society.

Finally, by discussing the controversial themes that lead to the division of positions in society, students can exercise their argumentative capacity since the narrated episode contemplates the social, political, and scientific values. This potential of the episode for the teaching of physics goes back to the epistemological theoretical matrix of Levinson (2006) to explore the SSI. It is possible to know and explore rational justifications from the perspective of ethics of cordial reason (Cortina, 2007) for the agreement or disagreement of political and social positions regarding gender issues, ethics in research, and pro-war decisions. How do we debate this historical narrative in favor of Education to promote PEACE and SJ (D´Ambrósio, 2006; and 2021)? Therefore, this provides the elaboration of arguments between students who are willing to communicate, bringing their own knowledge and values, to defend their own justifications, convince other people, or abandon what they already believed, in liberating education (Freire, 1998). Moreover, through this, they express their ways of thinking in an attempt to resolve disagreements present in the episode analyzed and discussed.

8.5 SOME CONSIDERATIONS

This chapter brought a proposal of a theoretical framework by merging the perspectives of D'Ambrósio, Freire, and Cortina aligned with the theoretical foundations of SSI to guide the historiography that supports dialogue with contemporary themes and fosters the promotion of SJ. This theoretical framework combined with theoretical feminist references was applied in the development of a historical narrative aiming to think about women in physics, which allowed at least thinking about the following:

- contextualized gender in science;
- the religious context of a scientist as an element of exclusion in scientific practices;
- the context of wars, their injustices, and consequences, which impose the need for exiles;
- the lack of scientific ethics that excludes researchers from publications;
- and war contexts as an application of physics knowledge to the service of political and economic interests that shape the world.

It is important to remember that many immigrants, refugees, and their descendants still face enormous racial and religious prejudices in different parts of the world.

Given the characterization of SSI, questioning elements of the trajectory of Lise Meitner brings inter- and transdisciplinary debates due to its own epistemic nature. That is, a typical problem of an SSI cannot be debated considering only STS aspects, but they should include the ethical, moral, SJ, and human rights promoter elements. It is not possible to think of the injustices suffered by Lise only from scientific publications. It is necessary to enter other broader aspects of her life into the socio-political context in which she lived. Her condition as a woman was a determining factor to exclude her from access to schooling, paid jobs, adequate workplaces, scientific publications, and her absence among Nobel nominees. Besides, her Jewish origin, even if she had converted to Protestantism, was the leading cause of her exile. So, similarly could be the justification for the association between nuclear fission theory and its possible application.

The episode also brings society's place in the face of the episode of the war and the people chosen to attribute glories and guilt. The atomic bomb explosion in 1945, as another element in the narrated episode, brings to light the recent historical moment between Russia and Ukraine, and other current bellicouse conflicts. Obviously, analyses of the relationship between STS, besides the motivations and geopolitical implications, should not be assessed only by the look of physics but certainly by an integrated view of the different areas of natural sciences, history, and philosophy of science and social sciences.

Using Meitner's academic trajectory, associating SSI with HP studies broadens the critique of the model of pure, rational, empiricist science called modern science. This criticism is already present in the HS community with the ideas of epistemologists such as Thomas Kuhn and Paul Feyerabend. However, their theories and demarcation criteria do not consider aspects related to the gender issue, for example. Including discussions about the gender identity of scientists, supported by a feminist theorization, allow demonstrating a concern with the model of production of science and white and cis-hetero hegemony among its characters.

Criticism is based on the need for science to meet the interests of the plurality of characters that compose it. Since the work of education must be in the direction of presenting a plural science (Aikenhead, 2006, 2010; and D'Ambrósio, 1985, 2006, 2021), and this plurality has not yet reached the desired level, it is necessary to criticize the reasons why a majority group of men still perpetuate themselves in spaces of science, especially spaces of power. From a pedagogical discussion centered on Meitner's trajectory, it is possible to promote debates and encourage the argumentative posture in the training of students. SSI can promote discussions about NOS considering the identity of those who make this science. Talking about NOS aspects in classrooms and not highlighting gender, racial-ethnic identification, and class is being oblivious to the diversity of characters in science. As Chadha (1998) tells us, as long as the category of gender is not important for discussions about science and its nature, it will not be possible to understand that the identities of the characters of this science are also responsible for the conception of the current model of science.

It advocates the need to rethink the concept of NOS to include in this debate the gender prejudice inherent in scientific practice. Going further, as long as gender, racial-ethnic groups, sexuality, generation, social class, and many other identity markers are not on the agenda for understanding NOS, it will not be possible to understand the effects of the diversity of scientists on its development.

In summary, this chapter proposes a theoretical framework that merges SSI literature proposals, HP contributions to science education, Paulo Freire's pedagogical perspective, D'Ambrósio's Ethnomathematics historiography, and Cortinas's cordial reason for the PER community considerations. It is argued that this theoretical framework can assist the choice of episodes and the aspects to emphasize in the development of historical narratives, committed to culturally responsive physics teaching. Theoretical and applied aspects presented above do not intend to be a rigid structure to be replicated. Instead, the argumentation intends to inspire researchers and teachers to think and analyze how some aspects emphasized in any episode can bring awareness to the 21st-centurýs concerns, and allow us to know and reflect on the past in order to understand and become aware of critical aspects of the present.

However, considering the engagement of the PER community in identifying the role of physics in the historical constitution of this model of social organization and exploitation of nature and human beings, the theoretical framework proposed here seems suitable for addressing the promotion of democracy and the formation of people committed to SJ. Research can select episodes of HP adequate to promote a decolonizing and anti-racist education, as well as develop historical narratives able to bring conciousness about devaluation and epistemic prejudices on Indigenous Original *Sapientia*. This is the first step to take actions to achieve reparation and reconciliation with so many colonized people around the globe. Once the historical narratives are able to bring SSI, the PER community can use the theoretical references they are acquainted with to develop pedagogical and didatica proposals, to engage students to learn physics while debate deforestation, environmental degradation, climate changing and their consequences, such as the environmental racism, inequalities, and the extreme natural phenomena around the world.

Besides, the didactic proposals prepared by the PER community can offer fundamentals to the perspective of practicing teachers, to educate people to a better and fair world.

ACKNOWLEDGMENTS

The first author is grateful for the support of FAPESP (Proc. No. 2016/05843-4).

REFERENCES

Aikenhead, G. S., Can. J. Sci. Math. Techno. Educ. **6**, 387–399 (2006).
Aikenhead, G. S. and Elliott, D., Can. J. Sci. Math. Techno. Educ. **10**, 321–338 (2010).
Allchin, D., Sci. Educ. **95**, 518–542 (2011).
Bagdonas, A. and Silva, C. C., Sci. Educ. **24**, 1173–1199 (2015).
Chowdhury, T. *et al.*, Sci. Educ. Int. **31**(2), 203–208 (2020).
Clarke, E. H., *Sex in Education; Or, A Fair Chance for Girls* (Riverside Press, Cambridge, 1884), see https://archive.org/details/sexineducationor00clariala/page/n5/mode/2up.
Cortina, A., Isegoría **37**, 113–126 (2007).
Crawford, E. *et al.*, Phys. Today **50**(9), 26–32 (2008).

D'Ambrósio, U., Revista História da Matemática para Professores **7**(1), 14–25 (2021).
D'Ambrósio, U., Learning Math. **5**(1), 44–48 (1985).
D'Ambrósio, U., *Ethnomathematics: Link Between Traditions and Modernity* (Brill, 2006).
Forato, T. C. M., *Teaching Science with Context Historical, Philosophical, and Sociological Approaches*, 1st ed., edited by M. E. B. Prestes and C. C. Silva. (Org.) (Springer, New York, 2018), pp. 293–311.
Freire, P., *Pedagogy of Freedom* (Roman and Littlefield, Maryland, 1998).
Freire, P., *Pedagogy of the Oppressed* (Herder and Herder, New York, 1970).
Gavroglu, K. *et al.*, History Sci. **46**(2), 153–175 (2008).
Hahn, O., Sci. Am. **198**, 76–84 (1958).
Hammerich, P., *The Nature of Science in Science Education: Rationales and Strategies*, edited by W. McComas (Kluwer Academic Publishers, Dordrecht, 2000), pp. 127–136.
Haraway, D., Fem. Stud. **14**(3), 575–599 (1988).
Harding, S., *Ciencia y Feminismo*, translated by edited by P. Manzano (Morata, Madrid, 1996).
Heering, P. and Höttecke, D., *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. Matthews (Springer, Dordrecht, 2014), pp. 1473–1502.
Hodson, D., Can. J. Sci. Math. Technol. Educ. **10**(3), 197–206 (2010).
Hodson, D., Can. J. Sci. Math. Techn. Educ. **20**, 592–622 (2020).
Ibrahim, S. *et al.*, J. Activist Sci. Technol. Educ. **12**(1), 33–52 (2022).
Keller, E. F., Cad. Pagu. **27**, 13–34 (2006).
KolstØ, S. D., Int. J. Sci. Educ. **23**(9), 877–901 (2001).
Levinson, R., Int. J. Sci. Educ. **28**(10), 1201–1224 (2006).
Lima, I. P. C., Ph.D. theses (Salvador: Facedufba, 2019), p. 181.
Macalalag, A. Z. *et al.*, Cult. Stud. Sci. Educ. **15**, 389–413 (2020).
Mårtensson-Pendrill, A., Phys. Educ. **41**, 493–501 (2006).
Matthews, M. R., *International Handbook of Research in History, Philosophy and Science Teaching* (Springer, Dordrecht, 2014).
Matthews, M., Sci. Educ. **1**, 11–47 (1992).
Meitner, L., Interview with Lise Meitner. [Interview with Archives for the History of Quantum Physics project] Carried out by Otto Frisch and Thomas Kuhn. Cambridge, England (1963).
Meitner, L., Nature **143**, 239–240 (1939).
Meitner, L., Phys. Today **13**(8), 16 (1960).
Metz, D. *et al.*, Sci. Educ. **16**, 313–334 (2007).
Millar, R., *Science Today: Problem or Crisis?*, edited by R. Levinson and J. Thomas (Routledge, London, 1997), pp. 87–101.
Oliveira, B. *et al.*, Transversal: Int. J. Hist. Sci. (5), 146–156 (2018).
Pingree, D., Isis **83**(4), 554–563 (1992).
Reiss, M., *Social Justice, Education and Identity*, edited by C. Vincent (Routledge Falmer, London, 2003), pp. 153–165.
Rife, P., *Lise Meitner and the Dawn of the Nuclear age* (Plunkett Lake Press, 2017).
Rudge, D. W. and Howe, E. M., Sci. Educ. **18**, 561–580 (2009).
Sadler, T. D. and Fowler, S., Sci. Educ. **90**, 986–1004 (2006).
Sadler, T. D. and Zeidler, D. L., Sci. Educ. **88**, 4–27 (2004).
Sadler, T. D., J. Res. Sci. Teach. **41**(5), 513–536 (2004).
Santos, W. L. P. D., Sci. Educ. **93**, 361–382 (2009).
Saunders, K. J. and Rennie, L. J., Res. Sci. Educ. **43**(1), 253–274 (2013).
Scheich, E., Osiris **12**, 143–168 (1997).
Schiebinger, L., *Has the Feminism Changed Science?* (Harvard University Press, Cambridge, 1999).
Sedeño, E. P. *et al.*, *Interacciones Ciencia y Género: Discursos y Prácticas Científicas de Mujeres*, edited by M. J. Barral (Içaria, Barcelona, 1999), pp. 17–37.
Shapin, S. and Schaffer, S., *Leviathan and the Air-Pump* (University Press, Princeton, 1985).
Sime, R. L., *Lise Meitner: A Life in Physics. Berkley and Los Angeles* (University of California Press, 1996).
Tessum, C. W. *et al.*, Proc. Natl. Acad. Sci. U.S.A. **116**(13), 6001–6006 (2019).
UNESCO. (2017). Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM), see http://creativecommons.org/licenses/by-sa/3.0/igo/.
Yap, S. F., Issues Educ. Res. **24**(3), 299–319 (2014).
Zeidler, D. L. and Nichols, B., J. Elem. Sci. Educ. **21**(2), 49–58 (2009).
Zeidler, D. L., *Handbook of Research on Science Education*, edited by N. G. Lederman and S. K. Abell (Routledge, New York, 2014), Vol. II, pp. 697–726.
Zeidler, D. L. *et al.*, Sci. Educ. **89**(3), 357–377 (2005).
Zeidler, D. L. *et al.*, Sci. Educ. **86**(3), 343–367 (2002).

CHAPTER 9

THE AIMS AND VALUES OF PHYSICS

Andreia Guerra and Ivã Gurgel

Guerra, A. and Gurgel, I., "The aims and values of physics," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 9-1–9-28.

9.1 INTRODUCTION

We live in a period where questions like "Why should we trust science?" and "What is the role of scientists in political decisions?" have been very present in society. These issues, which bring us back to History, Epistemology, and Social and Cultural Studies of Science, are also addressed nowadays by journalists. This comprehensive treatment of questions about the Nature of Science (NOS) occurs due to its harmony with speeches of groups and political agents that have raised them, a phenomenon that demonstrates that citizens have become concerned with them. As a result, science became popular in a way many did not expect.

Much of this debate has focused on specific themes. One of the most common cases is related to environmental issues. Recognition of the increase in the Earth's average temperature and the natural calamities that may occur due to different phenomena has required actions from agents at the highest political levels. For decades, we have seen the growing presence of environmental issues inside the political environment. If those issues were "flags" of the so-called "green parties," today they have become a topic that can define elections, demanding positions from any party. However, given the costs and economic impacts that involve the issue, influential leaders have preferred to deny scientific results, such as those produced by the IPCC (Intergovernmental Panel on Climate Change), to keep their agendas and supporters.

Although more restricted to biological sciences, another example deserves to be mentioned. The context of the pandemic "heated up" the debate regarding vaccines, closely related to the possibility of them causing autism in children (a correlation already proven to be unfounded). The rapid development of vaccines for Covid-19 raises suspicion among people who have little trust in science or scientists. However, the most intense controversy is about the option of governments forcing the population

https://doi.org/10.1063/9780735425514_009

The International Handbook of Physics Education Research: Special Topics, Mehmet Fatih Taşar and Paula R. L. Heron

to be vaccinated or creating forms of control such as "vaccination passports." In this case, through epidemiological analyses, should scientific results override a fundamental right and limit the freedom to come and go? People already used to act following science answer this question easily, but large groups have different attitudes. Today, we see countries struggling to overcome 70% of the vaccinated population.

In the political discussions regarding climate change and vaccination, it has become quite clear what interests and motivations are involved in denying science. However, it is not so evident in other cases. For example, a topic that has become dear to the physical sciences is the defense of the Flat Earth, as we have already accumulated centuries of evidence, including visual ones, of the Earth's sphericity. However, since the Cold War, we have kept a conspiracist imaginary about space exploration. "Theories" about humans having never set foot on the moon have been going on for decades. If in the past these speculations fueled the imagination of science fiction fans and those curious about life outside the Earth, today they are in line with those who believe that science is reduced to a weapon of political interest.

This chapter seeks to draw attention to the fact that understanding about NOS is related to questions and uneasiness present at a given time, mobilizing different issues for physics teaching, such as the aims and values of physics. We often imagine that an epistemological stance results from an attempt to define what knowledge is, whether expressing a more spontaneous response—as in the case of students—or constituting a systematic work—as in the cases of those who dedicate themselves to Epistemology as a philosophical discipline. However, it is crucial to notice that not only ordinary citizens but also philosophers and scientists have views of the sciences connected to the lived experiences of their time. It implies recognizing that they change throughout history, bearing the marks of time. We draw attention to this because it is common to essentialize philosophical positions, making them rigid labels.

Then, we understand that discussing the aims and values of Physics is to examine how different contexts allowed particular philosophical views to be constructed and how they mobilized proposals for the teaching of physics. Therefore, this chapter aims to discuss how the physics education research (PER) and physics teaching that works under the strand of History, Philosophy, and Sociology of Science (HPSS) have justified and presented proposals throughout its development, in dialogue with different philosophical views that discussed the aims and values of sciences, and, in particular, of physics. To fulfill this objective, we present some prominent Epistemology/Philosophy of Science and questions about science brought by Science Studies (SS). Then, we discuss how visions of science presented by and discussed inside Physics Teaching were created throughout history.

Physics was one of the first areas of knowledge to obtain recognition as a legitimate science. This made it a model science, which established (sometimes inappropriately) standards to be followed by other fields of knowledge. As a result, much of 20th-century epistemology was built based on physics, even though the statements often refer to science in general. Moreover, when we analyze the most influential authors of the period, such as Karl Popper and Thomas Kuhn, we see that they have physics as the basis of their training.

Physics is a science that seeks to understand natural phenomena on different scales of magnitude, from the subatomic world to the structure of the Universe as a whole. Many of the epistemological issues in the current philosophical literature result from this "boldness" of physics. For example, the clash between realist and anti-realist views currently takes place around the realism of unobservable entities, such as electrons, quarks, and others that make up the physical universe. Thus, discussing the objectives and values of physics is to analyze which epistemic practices allow us to understand these different scales of the Universe.

9.2 HISTORICAL REVIEW OF THE AIMS AND VALUES OF PHYSICS

9.2.1 From inductivism to logical positivism

The empirical-inductivist philosophical currents are one of the most cited in science education (SE) research. In this case, it is a negative reference, in the sense of indicating that this is a philosophical view to be avoided in physics teaching. As will be detailed in the next section, many researchers who analyzed student conceptions classified them as empirical-inductivists, and a lot of work was done discussing how to "overcome" this. However, we have to pay attention to the fact that the use of philosophical concepts to categorize student responses is better employed when considered as an approximation or free use of the term.

Empiricism, in general terms, consists of the defense that the foundation of knowledge is the observation of natural phenomena or, in its version of the modern period, data obtained through experimental practice. Inductivism would be the method by which one would start from observations or detailed data for the proposition of general laws. It should be done through a detailed observational protocol in which occurrences are evaluated under different experimental conditions.

We can summarize the empirical-inductivist assumptions as: there is objectivity in the observations made, and they can be repeated/reproduced; it is possible to establish sound procedures to delimit the occurrence or not of facts (i.e., it is possible to carry out a properly controlled experiment); it is possible to propose general laws based on observations and experiments; and finally, the inductive method makes it possible to differentiate science from metaphysics.

Inductivism and empiricism are philosophical doctrines that had significant influence between the 17th and early 20th centuries. Among the authors that can be considered as their representatives are Francis Bacon (1561–1626), David Hume (1711–1776), and John Stuart Mill (1806–1873), indicating three outstanding examples from different centuries. Scientists of great recognition, such as Isaac Newton (1642–1727) and André-Marie Ampère (1775–1836), also presented in some of their works positions that can be classified as typical of these currents.

In Novum Organum (1620), Bacon considers any knowledge that does not come from the inductive method applied to experimentation as premature abstractions. For him, in the path of true knowledge, “axioms are gradually elicited step by step so that we reach the most general axioms only at the very end; and the most general axioms come out not as notional, but as well defined” (Bacon, 2003, p. 17). Thus, all knowledge should be adequately grounded in data and observations.

Inductivist ideas remained defended for more than two hundred years. For example, Stuart Mill shows the idea that science is based on inferences from experimental data in his 1843 work *A System of Logic, Ratiocinative and Inductive*. For him, knowledge is obtained through the accumulation of data, when general propositions are obtained through the “collection of particulars” (Mill, 1981, p. 287).

Empiricists and inductivists view dialogue with the philosophy of the time in how the problem of knowledge validity is posed. The aims and values of physics that guide its practices are framed concerning how to obtain true knowledge. Different philosophical traditions have sought to find the essence of truth, that is, whether it resides in reason or some other “faculty of the soul,” in the experience or some modes of perception. Having found this “essence,” it would be necessary to delimit the procedure, the method which would guarantee that the correct way of proceeding would lead to a specific law or proposition. In summary, the question centers on the *genesis* of knowledge, demanding knowledge to be born in the correct way to generate knowledge. This way of posing the question about knowledge’s possibilities makes doctrines such as empirical-inductivism seek to eliminate the creative role of the human mind in constructing physics laws, theories, and experiments. Thus, Physics aims to reveal, from empirical studies without human intervention, laws that could explain natural phenomena.

Another factor marks the historical period of the epistemological views in question. The 18th century gave birth to a new place for sciences and physics. Academies and scientific societies supported experimental physics, valuing it and allowing public demonstrations of experiments to be carried out (Phillips, 2016). Universities, still tributary to their medieval heritage, valued productions that represented an abstract intellect due to the liberal arts that made up the *quadrivium*. Experimentation, the fruit of the mechanical arts, had little value inside those *places of knowledge*.

The foundation of the Royal Society in 1660 and the French Academy of Sciences in 1666, taking two striking examples, is a fundamental step toward establishing a new science, now called experimental, empirical, or inductive science. The Middle Ages had already witnessed the beginning of the growth of cities, which little by little led to economic and social changes. Together, new technologies were created making artisanal knowledge very sophisticated and generating a movement of mutual influence; practical knowledge needs theoretical knowledge and vice versa (Rossi, 2001). If this movement continues today, it is almost inevitable to place the industrial revolution as one of its central markers. Heat engines have become a synthesis of the relationship between science, technology, and society when thinking about their progressive aspect. Thus, experimental science carried out by practice, will be representative of this period, and epistemology will reflect the need to justify it.

The 19th century also saw the birth of another philosophical current, much criticized later, Positivism. Ian Hacking indicates that this philosophy can be delimited by six "instincts": (i) an emphasis upon verification, (ii) proobservation, (iii) anti-cause (there is no causality in nature), (iv) downplaying explanations, (v) anti-theoretical entities, and (vi) against metaphysics (Hacking, 2010, p. 41–42).

From the points raised by the author, the anti-metaphysical attitude is the one that synthesizes positivism. It is essential to note this philosophy's high level of demand, as it is not just a matter of preventing gods from being mobilized in scientific explanations. Likewise, reference should not be made to unobservable entities or explanatory causes. For example, in positivism, all Elementary Particle Physics must be considered a "sea" of speculation. But Positivism should not be regarded simply as naïve or unreasonable philosophy.

Auguste Comte (1798–1857) published several works in which he sought to characterize Positivism, among which his *Course of Positive Philosophy* (Comte, 2020) stands out, published between 1830 and 1842. Comte is a historicist, for whom the understanding of something passes through the understanding of its historical evolution. Through it, we can move towards progress, something well represented by his Three-State Law. Scientific/positive knowledge, based on reasoning and observation, would be the overcoming of metaphysical or abstract knowledge still attached to the search for the intimate causes of phenomena. It, in turn, is the overcoming of theological or fictitious knowledge when these causes refer to divine entities or of a similar nature.

Comte is the result of the period that Eric Hobsbawm (1996) defined as the "Age of Revolution (1789–1848)," a period that will continually give weight to the notion of progress. Positivist epistemology is born as a condition for formulating a project of society. The great demand for knowledge posed by positivism comes from the expectation that science is the lighthouse for society. Only through this would a more egalitarian society be reached (Fedi, 2017). Comte is considered one of the founders of Sociology for having defended the importance of a science of society. It is interesting to note that he calls Sociology Social Physics, showing the importance of Physics as a model for other sciences.

The turn of the 19th to the 20th century was described by historian NicolauSevcenko (2001) as a roller coaster. The western world came from progressive enthusiasm that made us believe that achieving everything we could imagine would be possible. There would be no limits to progress, and cities began to light up. In this way, the aims of physics were to progressively continue the works of the great physicists to get closer to the truth of the universe. And without metaphysics, physics is value-free. However, the culmination of this period was a war of unprecedented proportions that shook society in many ways. The post-war period, especially in German-speaking countries, came to be seen as a world guided by irrationality (Gay, 2001). In this context, a new positivism will rise in the 20th century.

One of the main institutional milestones of Philosophy of Science as an academic discipline was the creation of the Chair of History and Theory of Inductive Sciences at the University of Vienna in 1895, whose first professor was Ernst Mach (1838–1916) (Moulines, 2020). There, the concern was expressed in giving new contours to the relationship between Physics and Philosophy, in which the foundations of knowledge were well defined. A few years later, in 1922, Moritz Schlick (1882–1936) would occupy this chair and around him would start a scientific-philosophical movement known as the Vienna Circle.

The Vienna Circle sought to reconceptualize some of its antecedents (inductivism, empiricism, positivism) while maintaining some of its central concerns: an anti-metaphysical attitude, with the central issue of delimiting the appropriate method to justify scientific statements (Moulines, 2020). The emphasis on *justification* and not on the knowledge *genesis* is one of the distinctions between the Circle's philosophy of those of previous periods. It is more critical to substantiate the validity of given knowledge, regardless of how it was elaborated, than to guarantee its creation has taken place using a given method that legitimizes it. Later, Hans Reichenbach (1891–1953), a sympathizer of the circle and professor in Berlin, would make the distinction between "discovery context" and "justification context," with philosophical analysis to deal with the latter (Reichenbach, 1938).

The philosophies produced by members of the circle were called, among other terms, operationalism, logical empiricism, and logical positivism. However, one should not lose sight of the different views among the Circle participants. Between 1924 and 1936, a group of philosopher-scientists—most of them physicists and mathematicians—met periodically to discuss the foundations of science, and, in particular, of physics. In addition to Schlick, names such as Hans Hahn (1879–1934), Philipp Frank (1884–1966), Otto Neurath (1982–1945), Olga Hahn-Neurath (1882–1937), and Rudolf Carnap (1891–1970) participated (Uebel, 2021).

Driven by advances in the physical and formal sciences, different Circle participants sought Logic as a way to provide a good foundation for science. Here, Logic should be understood as an academic discipline that seeks to analyze how to relate different propositions and obtain consistent or inconsistent results. However, logic (or mathematics) would not be the foundation of science. Any area of knowledge should be based on verifiable propositions, that is, propositions that can be identified as true or false, taking the world as a reference. This verifiability principle was inspired by the work of the so-called first Wittgenstein, for whom the meaning of any utterance is none other than its verification conditions (Moulines, 2020).

Empirical propositions, also called protocol statements or observational statements, lay the groundwork for theoretical statements. The latter would guarantee consistency to a theory and allow the former to be, a posteriori, deduced from them. However, the authors' project did not sustain itself both due to the difficulty in limiting the meanings of the elements present in the observational statements to empirical correspondents and due to the difficulty in being reductionist in the sense that theoretical propositions are dependent on observational statements.

Nevertheless, the non-continuity of the Circle occurs for well-marked historical reasons. As indicated earlier, a kind of "irrationalist culture" was increasing in the post-World War I period, which motivated different thinkers to defend opposing positions. In 1929, Hahn, Neurath, and Carnap wrote a manifesto entitled "The Scientific Worldview: the Vienna Circle," arguing that the philosophy produced by them would bring about a new worldview, free from dogmatism and metaphysical confusion. Therefore, the aim of physics is to produce understanding about natural phenomena free from dogmatism and metaphysical confusion. The manifesto was a tribute to Schlick, without its authors knowing that he would be killed by Nazis a few years later. Other members of the Circle did not have such a tragic end, but most of them had to go into exile in other countries with the advance of Nazism.

9.2.2 The problem of scientific change and the historicist turn

The Epistemology of the 20th century had many contributions, and it is always an arduous task to synthesize them. For this chapter, we consider that commenting on authors who have become more popular is the best option, as they are the same ones that exerted the most influence on educational proposals aimed at NOS. One of the characteristics of the Philosophy of Science of the 20th century, which characterizes it as a different type of philosophical reflection compared to the theories of knowledge of previous centuries, is a gradual abandonment of the attempt to prescribe what science should be. Philosophers such as René Descartes (1596–1650) and Immanuel Kant (1724–1804) sought to create philosophical systems that, among other things, founded the possibilities of knowing the truth. From this, it could be prescribed how to produce knowledge of maximum validity. When modern science was taking shape, the best attitude would be to state how it should be. A new perspective was established in the 20th century, especially from 1930. In a context in which Physics, in particular, was advancing in an unprecedented way, the focus became to understand how science, and then physics occur. From the prescription about the aims and values in physics, one returns to the description; from trying to understand how physics should be, it turns to physics as it is. Answering the central question of the Epistemology of Science—*what is science?*—now becomes the result of a kind of empirical study, having science as the object.

Karl Popper (1902–1994) is often the first name mentioned when referring to post-positivist philosophy. Although he was not a member of the Vienna Circle, Popper debated his ideas with characters such as Carnap and Feigl, being described by Neurath as a loyal opposition to the Circle (Moulines, 2020). The demarcation problem, that is, the possibility of separating what is science from what is not, will be one of the main axes of Popper's epistemology. It was present in one of his first and best-known works, published in 1934, *Logik der Forschung* (The Logic of Research, which in the English edition of 1959 was called The Logic of Scientific Discovery).

In *Logik*, Popper presents a critique of inductivism. As important as observations are to produce knowledge, basing it on them, that is, betting all the chips on the possibility of being able to propose general theories from particular observations, is something invalid for Popper. Natural regularities or facts are always conjectural; new elements can be verified in new studies. This criticism will become known for a passage in which Popper exposes it with certain poetry:

> "Now it is far from obvious, from a logical point of view, that we are justified in inferring universal statements from singular ones, no matter how numerous; for any conclusion drawn in this way may always turn out to be false: no matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white." (Popper, 2002, p. 4)

One of the difficulties pointed out by Popper is related to the impossibility of having a clear criterion on how many observations would be necessary to generalize a given statement. For example, the lack of a certain “saturation point” would require an infinite number of observations. But the reference to “a logical point of view” in Popper’s phrase indicates another problem. More than one inference can be made from the same occurrence. For example, from observations of white swans, it is possible to say: “all swans are white” and “all swans are white or black.” Our spontaneous intuition would certainly stick with the first statement. But the second contains the first, giving it the same validity and even greater generality.

Popper’s epistemology is called critic—sometimes called rationalist or realist—for defending a distinction between scientific knowledge in relation to other forms of knowledge while knowing how to recognize its validity limits. For Popper, scientific theories, and physics theories never lose the status of conjecture, as they are never definitively proven. It is necessary to be careful with the term conjecture as it is easily interpreted as something that is proposed without great commitment to reality. It is certainly not what the author had in mind. What must be accepted is that even the best-justified knowledge is not definitive and can constantly be reformulated, thus recognizing the mutable character of science. However, Popper’s work comes to value one aspect valued in discussions about NOS teaching, the positive role of creativity and imagination in physics.

Still following the Popperian proposal, when an observation or experience in physics presents data that agrees with a theoretical proposition, we must say that it has been corroborated. The more correspondence one has with experience, the more a theory is confirmed, thus having differentiated validity. Nevertheless, this process does not allow us to claim that a hypothesis has been proven, which would make it crystallized truth. So, the aims of physics are to produce explanations of natural phenomena that could be counterfeit.

In the falsificationist method, physics is distinguished from metaphysics by committing to consistently producing falsifiable statements; that is, some empirical evidence may contradict it. The more falsifiable, the more scientific a physics theory should be considered. Although we can never prove them, the fact that we can discard propositions allows us to envision a convergence process to reality, allowing us to trust science as the best available knowledge.

Much discussion took place around Popper’s work, for example, about the possibility of corroboration of falsifications being reviewed, which would demonstrate that the author’s demarcation proposal would not be rigid. The recognition that the observations themselves are theory-laden, which would explain disagreements about the meaning of a test, also imposed limits on the Popperian proposal.[1]

---

[1] The theoretical dependence on observation is a topic addressed by Norman Hanson (1958).

Another author of great recognition who elaborates a work that brings counterpoints to Popper was Thomas Kuhn (1922–1996). His writings are from the second half of the 20th century, when the aims and values of physics changed. If science had been synonymous with progress at the beginning of the 20th century, its presence in the great wars through chemical weapons and the construction of bombs brought new insights into science and scientists. The warlike side of science revealed its non-neutral character, which could also be verified in scientific productions of industrial interest. Material production benefits society, but it is also linked to many problems, from exploiting workers to environmental issues. The difficulty in separating science, considered pure, from its biased uses has also led to new perspectives on how knowledge is produced.

Kuhn's best-known work is his 1962 book *The Structure of Scientific Revolutions* (Kuhn, 2012). Recognizing the two elements of Kuhn's work makes it easier to understand it. First, a view that emphasizes the collective character of science. To understand science, according to Kuhn, is to understand the scientific community not only in its institutional forms of organization but also highlighting how the fact that the production of knowledge is a work shared by scientists affects the characteristics of the knowledge produced. Another important pillar is the concern with the problem of "scientific change," which became central to the epistemology of the period (Laudan, 1978). The problem is to understand how science develops over time, considering the possibility of new ways of producing knowledge. In this case, it is not so simple to speak of progress in the sciences.

A term that will gain great prominence in Kuhn's work is the notion of paradigm. It characterizes normal science, that is, the science based on previous achievements. When we observe science focusing on individuals, we are easily led to think that their proposals result from the pure activity of their intellect, which would demonstrate their genius. However, looking at history with a readjusted focus, we see that each contribution is a tributary of the production of other groups of scientists.

In this way, physics is a complex activity in which the elaboration of an explanation for a given phenomenon is the result, even if unconscious, of a series of elements such as ontological and epistemological assumptions about the object of study, theoretical options that shape its understanding, adoption of experimental protocols, use of the analysis methods that allow the systematization of data, forms of data presentation and interpretation, and other elements that make up the physics activity. Due to this complexity, hardly a scientist would develop this set of elements alone. Hence, he or she takes advantage of all theoretical and practical knowledge available in his or her time. So, physics is not value-free.

The term paradigm is difficult to define, leaving it quite imprecise, something that has been widely criticized (Masterman, 1970). It represents how scientists act while attempting to explain a group of phenomena of the same nature. In a way, it is a pattern that is followed and defines a coherent set of philosophical assumptions, research questions, theoretical approaches, investigation procedures, modes of interpretation, and data validation criteria used by scientists. The most important point of the previous sentence is perhaps the term "coherent," which emphasizes that the paradigm is something very organic. Thus, theory, experiment, and other elements of physics practice are very closely linked. It may help us understand why the creation of a paradigm is a historical process and not the action of an individual or even something a group creates overnight. Thus, the way a scientist thinks or acts results from practices far beyond him.

Scientific activity done in normal science can involve a lot of creativity in solving “puzzles.” The investigations will be limited to what the paradigm defines as a legitimate problem to be solved. As they are immersed in and dependent on the paradigm, scientists would rarely abandon it. But when this process occurs, we have a scientific revolution. But it is worth noting that a revolution does not occur by chance. Still, according to Kuhn, accumulating anomalies (problems the paradigm should solve but does not) leads scientists to seek new paradigms. A scientific revolution is also not a process that takes place in short periods. It can take decades for a new paradigm to be properly articulated, becoming a new standard to be followed.

One aspect of Kuhn’s work that divides how his work received relates to theses that flirt with relativism. The paradigm can be seen as something that closes in on itself, radicalizing ideas such as theoretical dependence on observations. In a way, scientists would be conditioned to see what the paradigm allows. The paradigms’ incommensurability would indicate that the two paradigms do not refer to the same reality, suggesting that empirical evidence would be underdetermined. Thus, the paradigm shift could be reduced to a coercion process.

The relative prominence given to Popper and Kuhn is not due to the undeniable importance of their works, but because they were the ones who had the most capillarity in the formation of conceptions about NOS. The so-called “Popper x Kuhn debate” marked the epistemology of the period (Lakatos and Musgrave, 1970), influencing essential works. Imre Lakatos is an essential example of this. Coming from Hungary, a country where dialectical materialism prevailed, Lakatos will study Popper’s work in England and be an important follower. He seeks to historicize Popper’s work in a new way, treating the problem of scientific change from a critical rationalist perspective.

Lakatos (1989) developed the Methodology of Scientific Research Programs seeking to recognize a thread of rationality in choosing between theories, thus showing some scientific progress. Research Program guides the work of scientists and comprises a firm core to a protective belt. In the first one, in which there is a negative heuristic, there are primary hypotheses and concepts which characterize the Program that scientists are not willing to negotiate. In the second case, where there is a positive heuristic, there is room for negotiating the reformulation or readjustment of theoretical elements, giving flexibility to the Program. Lakatos’ proposal tries not to make science rigid to the point where a contradiction or empirical inadequacy already invalidates a theory. On the other hand, it seeks not to fall into relativism, in which irrational factors prevail in choosing a particular scientific approach.

Research Programs are transitory; that is, they have a limited existence in time. A Program is in a progressive phase when its theoretical elements interpret the empirical data obtained. The Program is considered regressive if theoretical advance lags behind experimental or observational novelties. The rationality of the historical process would be that scientists always prefer progressive programs, thus privileging scientific advancement. Besides the difference between Lakatos and Kuhn's arguments, we can summarize that in both philosophies of science, the aim of physics is not to develop theories or experimental inquiries to understand natural phenomena but to develop practices that allow understanding of natural phenomena considering that it is not possible to achieve a complete comprehension about them.

Paul Feyerabend is also a significant author in epistemology, known for his epistemological anarchism, which refuses any attempt to define the scientific method. He was one of the main interlocutors and opponents of Lakatos and regarding what has been exposed here, he indicates that there is some difficulty in "dimensioning" a Research Program. He argues that if we understand Physics as a Research Program, we will always see it as progressive. At the other extreme, if we define a new line of work (e.g., String Theory) as a Research Program, we will easily only see its regressive aspects, which should lead to its abandonment (something that normally does not happen). Thus, if Lakatos' epistemology would allow a look at history based on the problem of demarcation, in which one could recognize in scientific development elements that allowed its advance and separate them from externalist aspects (something Lakatos called rational reconstruction), Feyerabend's criticism places a limit on this proposal (unfortunately Lakatos' untimely death prevented him from responding to Feyerabend's criticisms) (Motterlini, 1999).

The authors mentioned in this section are the ones who had wide circulation in academic circles that privilege the English language, and consequently, their ideas are the most present in debates about NOS that fall within the same tradition. However, even if briefly, it is important to highlight other authors.

The French philosopher Gaston Bachelard (1884–1962) is very influential in Latin-speaking countries. His work began in the late 1920s and is the foundation of the French tradition in historical epistemology. The author analyzes how scientific thinking occurs over time, characterizing the development of science as a process in which thinking undergoes an overcoming to produce more complex concepts rationally. The epistemological obstacle concept explains why epistemological ruptures are rare in the History of Science and, in particular, in the History of Physics (Bachelard, 1993). Instead, current ways of thinking, the result of common sense or even already created in the science area, are consolidated in such a way that they become impediments to new mentalities; in Bachelard's terms, they become epistemological obstacles.

As a closure to this section, we mention the name of Ludwik Fleck (1896–1961). He published in 1935 the work *Genesis and Development of a Scientific Fact*. He proposed that understanding the creation of scientific ideas is to seek their social genesis, with knowledge being a production that involves a collective of people. The notion of thought style is mobilized to describe this collective way of thinking. The thought style is a "directed perception, with corresponding mental and objective assimilation of what has been so perceived" (Fleck, 1981, p. 99). The style indicates relevant problems to be solved and the best ways to do it.

One of the main aspects of Fleck's work is his discussion regarding the circulation of knowledge. Fleck distinguishes the esoteric circle—formed by specialists—from a wider one, the exoteric, which includes those interested in the thematic in question. The transit of ideas between these circles, both *intra*- and *interthought* collective, allows concepts to be rethought and can stimulate the creation of new perspectives on the world.

Despite the undeniable importance of epistemology produced in the second and third quarters of the 20th century (the period in which most important works were published, even if some authors had other publications after this temporal division), new perspectives on science emerged from the 1970s. As a result, there are various issues and themes which we will gather here around the Science Studies label.

9.2.3 Science studies

The adoption of the term Science Studies (SS) is associated with the attempt not to restrict analysis and reflections on science to the more traditional disciplines of humanities, such as History and Philosophy. SS presents itself as a strongly interdisciplinary area and is an extensive current research area. It seeks to understand sciences not only in their historical and epistemological aspects but also in their linguistic, semiotic, cultural, social, political, and legal aspects. It is common to find in SS departments or graduate programs researchers with initial training in the most different areas, from Physics to Philosophy, Architecture, Nursing, Psychology, Law, and many others.

Although it deals with questions about the nature of knowledge, SS distances itself from the more traditional Philosophy studies because they do not aim to compose a system that seeks to justify the foundation of true knowledge. The central question of Epistemology that shapes the aims and values of physics—*what is knowledge?*—and others derived from it, such as the problem of demarcation or scientific change, lose some of their leading roles. SS also distances itself from authors who already avoided prescribing how science should be. Although authors such as Bachelard, Kuhn, and Lakatos sought to analyze sciences and physics from how they were constituted, the result of their work is intended to present a model for how physics develops. This attempt is abandoned in SS (Pestre, 2006).

SS research tends to have a very narrow focus, being carried out around case studies. In the case of research in History, attempts to establish a single narrative line for history are abandoned, moving away from long-term historiography (*longue durée*). This "look with a magnifying glass" at episodes involving science aims to reveal all its contingencies. It seeks to understand all the characteristics—individual and collective—that mark this knowledge as a human production. Recognizing this human mark that cannot be eliminated from science means identifying our virtues in the knowledge produced and recognizing our limits and defects. As Dominique Pestre defines it:

> "The studies on science and science practices that have made history in the last decades have denaturalized the object "science," they have de-essentialized it, de-idealized it. It is postulated that there is no evidence that the science object exists identical to itself over time, and that its identity is unproblematic" (Pestre, 2006, p. 6, translated by authors).

Research in SS has authors from different schools, such as Harry Collins and Bruno Latour, some of them with their first works in the 1970s and 1980s. This is the case of David Bloor, one of the founders of the "Edinburgh School," also known as "Strong Program of the Sociology of Scientific Knowledge." In 1976, he published "Knowledge and Social Imagery" in which there are some central SS characteristics. The author claims that the debates on the nature of knowledge are not the responsibility of Philosophy alone, and Sociology should deal with them. The latter had only dealt with scientific institutions and could go further, analyzing the cognitive content of science itself.

We see Bloor's break with the then-current philosophical tradition in the very definition of his object of study. For him, knowledge is not just true belief but "knowledge for the sociologist is whatever people take to be knowledge" (Bloor, 1991, p. 5). Thus, understanding why specific knowledge is legitimized becomes an essential element. However, it is his four programmatic principles that will mark his work: (i) Causality: all statements must be analyzed in historical, intellectual, institutional, social, and cultural contexts; (ii) Impartiality: avoiding judgments about true and false, rational and irrational, etc.; (iii) Symmetry: following different actors who participated in science in the same way; and (iv) Reflexivity: applying the rules to the explanations given in historical studies.

If the programmatic nature of Bloor's work allowed a project of radical historicization of scientific knowledge to be disseminated, the somewhat rigid profile of its principles became its Achilles heel. For example, if the impartiality principle stands out for its supposed neutrality, the causality principle, which is interesting at first, proves complicated by placing the social to be external to knowledge to be its explanatory cause. However, this did not invalidate his proposal but showed the need to mature some aspects.

One of the aspects that marks SS authors and generates much controversy is its approach to anti-realist theses. It is necessary to emphasize that anti-realism is not related to anti-science movements. It is more a question of skepticism towards unobservable entities such as elementary physics particles. Recognizing the limits of science and the validity of non-Western traditional knowledge does not mean to give the same epistemological status to all interpretations of the natural world. It is crucial to avoid what is usually called scientism, that is, a blind and unappreciated belief in science.

Despite controversies regarding the quality of knowledge seen as the result of scientific activity, SS increasingly turns to the process of scientific production. Furthermore, SS highlighted the social character of science to point out that the development of scientific knowledge is not value-free. Thus, it is impossible to separate science from its environment. Science is then recognized as situated knowledge, which substantially depends on the local circumstances, people, epistemes, and politics of the place it develops. With that, many questions about science were put in check.

Andrew Pickering (1948–) highlights that science is performative and built based on transformative actions of material and human agencies, pointing to their temporality and contingency (Pickering, 1995). In this way, Pickering follows other SS scholars, such as Bruno Latour. However, the author does not consider that material and human agency are symmetrical, as required by the actor-network theory, defending human intentionality in scientific work. However, intentionality is understood as temporal and transformative, establishing itself in the encounter with the material agency.

Understanding how physics is practiced in laboratories is one of the central themes of SS. Discussions on the material agency highlight both instruments and other materials and the actions developed around them. The laboratory is not considered a space that validates scientific theories but a place where people and materials meet, and the practices established there are dynamic and local. Extrapolating the laboratory issue, scientific practices are understood as historical, local, temporal, and contingent undertakings, encompassing cultural performances and activities.

Recognition of material agency has brought attention to material culture, so experimental materials and procedures have become crucial to studying building physics. It means analyzing how instruments and experiments are constructed and reconstructed, the materials used in these processes, and the written or unwritten rules for their manipulation, in addition to studying the scientists and other social actors who developed or participated in these performances.

The understanding that science is imbricated in the social and historical context led to the consideration that "science and their societies co-produce and co-constitute each" (Harding, 2014, p. 53). Concepts considered as the science demarcation, considered value-free, such as objectivity, are problematized.

Objectivity is, by authors such as Sandra Harding (1935–), a historical concept, and therefore, it is transformed in response to different social and cultural processes. The author argues that defending objectivity as a universal epistemological category to qualify good science was a political project. In this way, Harding (2014) points out: "Its insistence that the objectivity of science depends on the value-freedom of its methods and results of research was itself simultaneously a commitment to a philosophy of science free of fascist and Cold War politics, and a specific political response to political threats against its adherents" (p. 154).

Historians of science, such as Daston and Galison (2007), developed a historical study considering that one cannot discuss objectivity without considering its binary: subjectivity. The authors discuss how self-scientists and objectivities have changed throughout history, taking illustrations as a primary source. They point out, for example, that in the 18th century, natural philosophy was the person considered capable of leading the adequate representation of a phenomenon or the object studied. Good illustrations were those that presented what the natural philosophy considered the essence of the phenomenon studied. Therefore, the natural philosophy should lead the illustrator to make the image to ensure the essence and not the details would be represented (Daston and Galison, 2007). In the 19th century, portraying a plant or a phenomenon was considered a scientific vice. In this context, mechanical objectivity in the sciences was present with the search for knowledge without traces of the researcher. The adequate scientific self was the one who distanced himself from the object studied because he had the pretension of knowledge not marked by prejudices or abilities, fantasies or judgment, desires or searches. Thus, the authors highlight changes in objectivity patterns and in scientists' preferred ways of observing natural phenomena, which are, in a way, related to changes in research technologies. Daston and Galison (2007) indicate that these different scientific selves do not constitute a succession history. There was no revolution to replace one with the other. Instead, their histories are of certain coexistence and permanence but with innovation. It points to epistemological pluralism in scientific development.

The discussion on the historicity of the objectivity concept and the self-scientist is taken up by feminist researchers such as Harding (2014) to defend "that such a new methodological strategy developed by the social justice movements as, for example, 'starting off research from the daily lives of economically and politically vulnerable groups,' as standpoint theory recommends, is itself an increasingly recognized new way to do a maximally reliable observation of natural and social relations. It is a new 'logic' or 'technology' of good research. It is a new methodology of 'right sight' that enables us to see aspects of natural and social phenomena that otherwise would be difficult or impossible to get into focus" (p. 161). These looks at science bring up new relationships between science and social justice because science is not perceived as linked to social justice issues only when offering final products to society but also in the possibility of incorporating these issues into its own doing.

In dialogue with SS, current historiographies of science avoid locating prestigious authors and analyzing their publications, which obtained greater recognition. HS writing has focused on scientific practices and knowledge circulation processes (Nyhart, 2016). It seeks to understand all dimensions and constraints of scientific work by analyzing elements such as the material culture involved in the experiments carried out, the creation process of representation for the analyzed phenomena, the role of different agents, from laboratory technicians to book illustrators, in the production of knowledge, and different ways of publication and dissemination of knowledge. In short, when we see science as culture, everything that can constitute a way of being in the world becomes an investigation element. The aim of physics is to develop practices that make sense in the culture in which we live to pose questions and answers about phenomena/environment. In this way, the values of physics are given by society where it is developing. In the following section, we will discuss how the PER and physics teaching under the strand of HPSS have been justified and presented proposals throughout its development, in dialogue with the philosophical views previously presented above.

9.3 THE CONSTRUCTION OF CONSENSUAL VIEWS OF THE NATURE OF SCIENCE

This section discusses proposals for implementing historical-philosophical approaches in physics teaching in dialogue with previous considerations and the aims and values attributed to such approaches. Defenses for introducing HS into physics teaching are not new. In the 1960s, for example, the Harvard Project of Physics (HPP) was produced and considered HS a fundamental pedagogical tool for developing curriculum and didactic materials for teaching physics in "senior high school."

The project was built in the context of the Cold War, in which actions were promoted to expand the population's level of scientific knowledge and attract young people to science.

In this scenario, the number of students who chose to study physics at the "senior high school" was declining. Gerald Holton (1969), one of the leaders of the HPP team, considered it crucial to attract students with different backgrounds and professional perspectives to the discipline to reverse this situation. After all, physics, as it influences everyone's lives, was necessary to prepare future workers both for the world of work and for living their own time (Holton, 1969). Based on that, he advocated a curriculum aiming to promote understanding of the humanistic aspects of physics: 'the sweeping power of a few fundamental laws, the use, and limit of models and mathematical formulations, the persistence of great themes, such as atomism, in the face of continual disproof of older models, the beautiful and sometimes remarkable story of how real people made physics (Holton, 1969, p. 21). In summary, the students must comprehend the physics characteristics that made it central to society. For this, a sequence of ideas capable of providing a more comprehensive view of explanations of the functioning of nature offered by physics should be presented (Holton, 1969).

The project construction, application, and evaluation had a group of collaborators from different areas such as physics, astronomy, chemistry, history, philosophy of science, and science teachers. Thus, in the didactic materials, there are proposals for replicating historical experiments, such as Galileo's inclined planes, discussion of the limits of some historical models to explain current issues, as the limits of the Greeks' explanations for the Earth movement, and discussion of the works of scientists. The social character of physics is referenced, considering that its development has implications in society, and society influences it. For example, thermodynamics development is presented in the context of the Industrial Revolution.

These highlight point to the project's historic character. Understanding the humanistic aspect of the nature of physics means recognizing that natural philosophers and scientists—real people—were responsible for the development of this science, which was influenced by society in the same way it changed society. Other educational contexts translated and applied the project. This acceptance might be explained by the fact that the project was in line with the historic turn in the philosophy of science discussed in the previous section. Thus, it, in opposing logical positivism, pointed to a historicist perspective highlighting physics as a human and social enterprise.

If, in the 1960s, we had the construction of a project that took HS as a didactic tool to reach the intended objective, an international conference on the history, philosophy, and teaching of science took place in 1989 in Tallahassee, USA. The International History, Philosophy, and Science Teaching (IHPST) Group emerged from this conference. The activities developed by the group played a crucial role in developing Physics Education Research (PER) that considers HS approaches. The second IHPST conference took place in 1992 in Kingston, Canada, and the group created the journal *Science & Education*, whose first editor was the philosopher of education Michael Matthews. The IHPST conferences then became biannual and, together with the journal, contributed to the consolidation of the area.

To slightly follow the path of such consolidation of the area, we will highlight the editorial of the first issue of *Science & Education* written by Michael Matthews. First, Matthews defends the HS in science teaching to reduce the artificial barrier between humanities and science, presenting it as an enterprise built by scientists who are not geniuses but real people with different personalities. In addition, he points out that HS can bring out the relationship of science with technology, industry, commerce, religion, and ethical values. The editor also refers to a document that recommends the inclusion of the philosophy of science's concepts in teaching to describe the nature of the scientific enterprise and its origins and limits in science teaching. Afterwards, Matthews (1992) points out that science teaching can be better conducted by promoting these understandings, and the interest and richness of science can also be better presented in the classroom.

It is worth noting that Matthews recognizes the diversity in HPSS in the 1992 editorial, pointing to the lack of clear consensus among scholars in these areas. However, considering that simplification is inherent to teaching, the author emphasizes that historians and philosophers of science can help build the simplification necessary for the historical-philosophical approach to make sense in education because they are knowledgeable about the issues in dispute in their fields. Thus, he calls for these researchers to contribute to the journal, pointing to the importance of looking at the History and Philosophy of Science fields to construct science curricula.

Many of the editorial arguments appear in-depth in Matthews's 1994 book "Science Teaching—The role of History and Philosophy of Science." Starting from the defense that science teaching should not be training in science but teaching about science, the author indicates that a historical approach is necessary. In addition to enabling a non-dogmatic and non-scientist look at science, HS has an intrinsic value since episodes of science and culture that should be familiar to all students, such as the Copernican Revolution, Gravitational theory, and thermodynamics, are better treated from a historical perspective. As in the 1992 editorial, Matthews points out that HS can humanize science by understanding the life and time of scientists, showing it as a human and social enterprise. He argues that, by doing this way, science teaching is no longer abstract and gains meaning for students.

We find in Matthews the defense of the history and philosophy of science in teaching with a descriptive character about what science is. By doing so, the discussions on issues raised by philosophers of science in the second half of the 20th-century guide Matthews's arguments. In opposing logical positivism, he approaches those who highlight science as knowledge constructed by scientists, who are men and women working in a collectivity of scientists. Matthews's arguments about the historical-philosophical approach in science teaching were adopted in PER.

Back to Matthews' arguments for historical approaches to teaching, it is essential to point out that such approaches help by combating students' spontaneous conceptions about physics content in science learning. In this way, HS presents itself with two main goals: to promote understanding of how physics has developed as a social and human enterprise and to be an instrument to teach great physics themes that everyone should know.

During the 1990s, discussions about the aims and values of the HS in physics teaching continued to be present in literature from the physics education area. In an increasingly systematic way, we find the association of HS with the teaching of NOS, which makes us dedicate attention to works that focus on this relationship. NOS is considered, mainly at the end of the 1990s, by the literature in the area and different national curricula, one of physics teaching's goals. HS is not understood as identical to NOS but as an important strategy for their teaching (McComas, 2008). Thus, not everyone who advocates and seeks ways to teach NOS will consider this contribution.[2]

A recurring argument about NOS teaching is that through its learning, the students can understand how science works, how scientists operate as a social group, and how society directs and is affected by the physics enterprise. This understanding enables students to manage the technological objects and processes they encounter in their daily lives, so that they can make decisions on socio-scientific issues and understand science as the main element of our culture (Drive *et al.*, 1996; McComas *et al.*, 1998; and Allchin, 2013).

We find in this context many studies on the views of students and teachers about NOS that point to their understanding of physics production and its products as positivist, empiricist-inductivist, ahistorical, individual, and decontextualized (Gil *et al.*, 2001). It is also discussed how such naïve views interfere not only in the students' view of physics but also in the learning of physics itself, and with that teaching NOS and HS are seen as ways to overcome such views.

Along with the defense of teaching NOS, we find arguments in favor of adequate simplifications for teaching the subject. In other words, as in Matthews (1992), it is recognized that HPSS is a disputed field and there are different views on how science works and how knowledge is established in science. However, it is argued that this should not prevent the search for ways to teach NOS. For example, McComas and Olson (1998) argue that teaching NOS based on disputes in the HPSS field is meaningless, and therefore, paths must be built to simplify this teaching.

Searching for possibilities, McComas and Olson (1998) analyze curriculum documents from English-speaking countries regarding science teaching and identify the elements of NOS that appear in the largest number of curricula. Based on these results, they build a list of statements about science, scientific practice, and scientists they consider consensual, therefore capable of being adopted to teach NOS.

Still, in the search for adequate simplifications, Osborne *et al.* (2003) conducted a Delphi study with 25 experts, science teachers, philosophers, sociologists of science, science disseminators, and scientists and presented a list of consensual statements about NOS among those consulted. The authors compared this list with the one indicated by McComas and Olson (1998) and identified convergence points

---

[2] It is important to note here that Matthews (2012) identifies HPS with NOS, stating that NOS is a new terminology for HPS. However, this is not the most ordinary understanding.

between them. Among the highlighted topics, we find that science is a human enterprise, as evidenced by how science has historically affected and been affected by social demands and expectations.

At the beginning of the 21st century, it strengthened the defense that it was possible to construct a list of statements about NOS that would respond to the desired simplification. Thus, a theoretical model of what to teach about NOS is built, known as consensus views (CV). The CV list composed underwent additions and changes to the one proposed by McComas and Olson (1998), but some general aspects can be highlighted. In summary, the CV points out that regardless of the assumed historical, philosophical or sociological current of science, there are statements about NOS that everyone agrees with and therefore should be taught in primary education and be part of educational assessments (McComas and Kamporaukis, 2015).

CV has become the most cited theoretical model of NOS in the literature in the area, generating a significant volume of empirical research and guiding proposals for evaluation instruments around the teaching of NOS (Erduran and Dagher, 2014). For example, based on it, Lederman *et al.* (2002) created a questionnaire, the V-NOS, with essay questions to verify students' understanding of NOS, that is, the tenets from the consensual list. Different versions of the V-NOS questionnaire have emerged, translated and applied outside the USA. The questionnaires to understand students' conceptions of NOS or assess the extent to which specific teaching methodologies allowed students to better understand the tenets presented derived much empirical research.

Still on CV, Abd-El-Khalic and Lederman (2000), for example, present research in which HS is used to teach NOS. They conclude that for HS to have some value, it is necessary to discuss NOS statements with students explicitly. In other words, when presenting the historical episode to students, the teacher must explicitly highlight the points from the consensual list that wishes to teach, such as the role of experiment in the construction of the studied theory or the creative power of scientists. These results had significant repercussions on PER, and much research has been developed aiming to select historical episodes to be worked on in the classroom, associating statements from the consensual list capable of being explained with the historical study.

For instance, McComas and Kampourakis (2015) present examples to teach aspects of CV from historical episodes in physics, geology, biology, and chemistry. Among the highlights, we find examples of physics to illustrate the humanistic character of science, understood as the creative aspect of science, its subjective component, and the historical, cultural, political, and social influences on science. Here it is essential to pay attention that when analyzing episodes taken by the authors, we find examples of scientists who looked at the same data and interpreted them differently due to their expectations and previous experiences. There are also examples of physicists who were creative in building methods of analysis and obtained inspirations that guided them from facts to conclusions. Even episodes presenting the development of science not only depend on scientists having new ideas and insights but also on their personalities. Other episodes highlight the exchange between scientists and the social context in which they worked. Thus, we find in the examples the names of Galileo Galilei, Tycho Brahe, Johannes Kepler among other characters who had notoriety in science, even if the recognition was late or their theories were questioned.

The examples that illustrate how the physicists constructed theories and experiments, with which scientists dialogued, how the scientific community received their works, what aspects of the social context they lived influenced their work, etc., illustrates the humanistic character of physics. In short, the humanistic character is emphasized to indicate that physics is not a solo work. Instead, the scientists operate as a social group immersed in society, and the scientific community and society influence them. Thus, we understand that the humanistic aspect highlighted in the CV aligns with Matthews (1992, 1994) and with the historicist perspective highlighted by 20th-century philosophers of science, as a counterpoint to logical positivism. In summary, HS is understood as a didactic tool to exemplify physics statements considered crucial for teaching.

This understanding of HS as a didactic tool allied to the concern with students' inadequate views of science find resonance in PER that points to concerns regarding how historical approaches are presented to students, that is, whether they represent history without anachronisms, whig history, or pseudoscience. For example, Forato *et al.* (2012) explain that it is crucial to follow the precepts of modern HS historiography, "providing well-grounded and faithful accounts" (p. 658). Based on results from empirical research developed in K-12 physics classes, the authors presented parameters to be followed so that the historical approach to teaching can allow well-informed teaching of physics (and physics itself) without errors and undue simplifications.

The CV movement grew in the first decade of the 21st century and drew criticism. For example, Matthews (2012) recognizes the contribution of the list to science teaching by putting NOS in the classroom, allowing teachers and students to stay informed and discuss some NOS topics, apart from the fact that CV provided an instrument to measure the learning of the subject. However, he identifies that CV, from what he calls the Lederman group, needs to be better articulated in the history and philosophy of science to resolve internal inconsistencies.

For example, regarding the list's statement that science has an empirical character, Matthews points out that without historical-philosophical support, this statement hides the role of idealization and abstraction in science, thus generating ambiguities. By not linking the claim that science has an empirical character with the philosophical discussion between realists and empiricists, constructivists, and instrumentalists, it is made clear that the world is real but not whether entities such as genes and atoms are also real. Matthews (2012) then highlights some statements by Lederman's group regarding theoretical entities, indicating that they seem to be affiliated with a constructivist philosophical view. Relying on Mario Bunge, the author argues that the problem is not with the supposed affiliation but in the absence of an explicit philosophical debate on realism that could lead to the conclusion that there are no disputes around the issue. Matthews (2012) points out other ambiguous statements and cites Lakatos to emphasize that the philosophy of science without the history of science is blind, as is the history of science without the philosophy of science. Considering Wittgenstein's terminology discussed by Irzik and Nola (2011), Matthews argues that it is helpful to understand NOS as a family resemblance of characteristics that justify different endeavors being called scientific. With this, he identifies NOS with the history and philosophy of science. He proposes replacing the teaching of NOS with the teaching of Features of Science (FOS) because then the emphasis would be on characteristics that would be discussed (rather than learned) historical and philosophical contributions.

Matthews (2012) highlights, in addition to the characteristics he recognizes present in the CV, empirical basis, scientific laws and theories, creativity, theoretical dependence, cultural immersion, scientific methods, and attempt, plus eleven other characteristics, including feminism. Regarding FOS, the author emphasizes that they should not be presented as dogmas but as subtle statements, and to recognize this subtlety, a historical-philosophical contribution is needed. Thus, he declares that a historical look leading students to identify the work and lives of characters who did science, such as Galileo, Newton, and others, is inevitable. Therefore, we can infer that Matthews (2012) reinforces the arguments of his 1994 book to argue that HS in teaching humanize science by making it possible to understand scientists' life and time.

Other criticisms of the consensus list point to the fact that CV claims can simplify the role and nature of observations and theories by virtually ignoring the role of models in science. When discussing the teaching of NOS in teacher education, Adúriz-Bravo and Izquierdo-Aymerich (2009) indicates what he calls the key ideas for teaching the subject, such as the fact of a similar relationship between a scientific model and the system it represents. In addition, CV's simplifications and ambiguities are pointed out as capable of generating an inadequate view of science. For example, the claim that science is tentative and that there is no method can lead to relativistic views of science.

In this regard, Martins (2015) discusses the exacerbated relativism that CV can provide. Regarding the statement "Scientific knowledge is tentative, durable and self-correcting" and that "errors" will be discovered and corrected, the author emphasizes that it suggests that error is something inevitable and not something that "can happen." Apart from that, the statement that knowledge is self-correcting can lead to the understanding that scientists usually can discover and correct their mistakes. Another problematic statement is that science has a subjective element. The author understands that this statement is related to the fact that observations are full of theory, which he agrees with. However, he emphasizes that knowledge is produced collectively and in dialogue, which gives it an intersubjective and not subjective character. Thus, for the author, the emphasis on subjectivity leads to the commonsense idea that theory is opinion. He recognizes that the distinction between law and theory goes against this commonsense view, but the isolated statements presented by CV end up giving rise to misunderstandings. Martins (2015) indicates that the CV was produced in opposition to a positivist, naïve realist, and commonsense view of science.

Relying on authors such as Matthews, who defends an image of moderate realist and rationalist science, Martins (2015) argues that relativism, "insofar as it makes human knowledge depend on historical, political, economic, social factors, on the of the time, etc., it should come—at least from the point of view of scientific education—accompanied by a discussion of the extent to which nature imposes restrictions on constructed knowledge" (p. 715). Without proposing to discard or disregard the CV, Martins (2015) is inspired by Clough (2007) and presents NOS to be worked on based on themes and issues. Supported by Driver *et al.* (1996), he argues that themes and issues are divided into two main axes: epistemological, sociological, and historical, to be treated in an interconnected way. For the author, "the properly epistemic aspects that characterize 'nature' of the produced knowledge come from a construction that is collective (intersubjective), historical and social." (p. 718). In the sense of discussing how scientists' individuality and the intersubjectivity of the social group of scientists talk about doing science.

Irzik and Nola (2011) also point out problems with CV. For example, emphasizing that knowledge is subjective and marked by theory raises questions: is scientific objectivity impossible? If not, why? Likewise, when we read that science is influenced by cultural and social issues, how can we explain that science produces reliable knowledge in distinct cultures and societies? Is society's influence on science good or bad? Is it possible to detect the influence of hostile societies and eliminate them? For the authors, these questions raised from the CV need to be answered if the goal is that students do not have a superficial understanding of NOS.

The authors also emphasize that the CV leaves out issues about the scientific investigation, which excludes teaching fundamental points about scientific practice, such as data collection, analysis, classification, and making inferences. They call attention to the monolithic image of science linked to the list since differences between disciplines and changes in scientific practice throughout history are disregarded. In this way, they highlighted the importance of considering the specificity of physics in the discussion of teaching NOS in PER.

They indicate teaching NOS based on the family resemblance, following Wittgenstein's understanding, to solve the problems raised. To this end, the focus should be on what is similar and different between the scientific disciplines from the following categories: activities, aims and values, methodologies and methodological rules, and products. According to the authors, based on these categories, it is possible to understand scientists as human beings driven by all kinds of non-cognitive goals when doing science, such as curiosity, fame, and engagement in improving society. However, unlike the CV case, it is understood that the critical inter-subjective process characteristic of scientific work makes "scientific discoveries" correct regardless of individual, social and cultural variations. Therefore, students can realize that disagreements between scientists are an integral part of science but are often rational and not purely subjective and arbitrary. Finally, the authors point out that the family resemblance approach (FRA) is also attractive because it is philosophically neutral once it is free from philosophical commitments such as realism, positivism, empiricism, constructivism, and others.

The path traced so far points out that in response to logical positivism, the teaching of NOS is intended to be descriptive and non-prescriptive, capable of highlighting physics as done by real men and women, and therefore not neutral. Furthermore, the historical approach to support this teaching must, following historical rigor, emphasize the work of natural philosophers and scientists, men and women, who cooperate as social actors in building science. The scientists' intersubjectivity ensures that scientific knowledge has objectivity and is not subject to the idiosyncrasies of scientists as individuals.

It is essential to underline those criticisms of CV that arise in a context where the HFSS field presents tensioning issues, such as those derived from SS.

9.4 NEW PERSPECTIVES FOR THE ROLE OF HISTORY, PHILOSOPHY AND SOCIAL STUDIES OF SCIENCE IN PHYSICS TEACHING

Issues raised by SS and new historiographical perspectives brought new views about science for those who work on the border between HS and physics teaching and those who take NOS as an axis of PER. In this context, to discuss the aims and values of physics is to teach not only epistemological aspects of the discipline but also cultural, social, political, and ethical aspects.

Allchin (2013) refers to SS to propose a teaching of NOS that moves away from the consensual list and leads students as citizens and consumers in contemporary society to assess the reliability of scientific statements relevant to the decision-making process of subjects with individual and social nature. In this way, the author defends a functional perspective for the teaching of NOS and points out that CV is not enough for this, as it turns to declarative teaching about science without aiming at functional actions in that learning. Still, on the CV, the author highlights that many of its items are irrelevant, and others that would be fundamental for the formation of students are not present in the list. As an example, he cites aspects related to the reliability of science and the social interaction of scientists, especially the item that refers to mutual criticism that provides science credibility.

From the defense that science is a social practice and not exclusively cognitive, Allchin (2013) emphasizes that SS contributes to understanding this aspect of science, but it is absent from science education. Based on the argument that the teacher's perspective and practical dimension of the educational context are fundamental to NOS teaching, the author indicates that science is a cognitive practice and a social one, referring to Bruno Latour. Considering the great challenge for teachers as making science meaningful for students, he proposes a specific kind of teaching NOS that he calls Whole Science, which feels all dimensions of reliability in scientific practice. He argues that science's social, conceptual, and experimental character should be treated together. Therefore, the material, cognitive and cultural aspects should be discussed in an integrated way. From this premise, he points to HS to teach NOS.

Allchin (2013) presents nine ways in which HS helps science teaching. First, the HS indicates science as a human enterprise, conducted by real people, fueled by curiosity and the desire to improve the human condition. Second, HS has the potential to assign names, temporality, and spatial dimension to ideas that often seem coldly objective and impersonal, leading to the appreciation of elegant ideas or experiments. It can also contribute to the understanding that science is practiced in scientific communities, highlighting the human element inherent to scientific practice without losing sight of scientific conclusions' general effectiveness and reliability. He thus reinforces a recurring argument that the human character of science brings students closer to science.

In addition, the author—referencing SS authors—calls for greater diversity in people represented in historical narratives and for more balanced perspectives, noting that historical models selected or emphasized by teachers are likely to affect those who seek a career in science. Still, it indicates that if the objective is to diversify the representation of who does science, the teachers should be concerned about highlighting many categories, such as economic class, personality type, or thinking style (mathematical vs verbal, abstract vs concrete, speculative vs conservative). HS can also help to understand diversity by presenting narratives that abandon reductionist Eurocentric visions for the development of science and teach about scientific contributions and forms of science in China, India, Africa, pre-Columbian America, Australia, Asia, and the Pacific.

In approaching SS, Allchin (2013) points to a science built by scientists who work collectively and do not follow white European man's stereotypes. Allchin (2013) suggests some historical examples, such as Galileo's judgment, which highlighted questions about the social and collective character of scientists' works, that is, the exchange and criticism between participants in the scientific community.

Erduran and Dagher (2014) also refer to SS to propose teaching NOS away from CV. The authors point out the recent emergence of SS in science education, indicating this new trend should be followed. Therefore, the scope of NOS should be expanded by turning to an interdisciplinary perspective that will make it possible to move from the narrow focus on the logic of the conceptual processes and scientific results in favor of a perspective of practices in which scientists engage. Based on this, they present a proposal called the Family Resemblance Approach (FRA). The proposal starts from the framework proposed by Irzik and Nola. It intends to be broad enough to accommodate a variety of aspects of science, including epistemic, cognitive, and social aspects. Irzik and Nola's framework has been extended in significant ways by the authors to incorporate more explicitly the social aspects of science.

Questions about science brought by SS mobilized changes in the HS field, so that, as pointed out by Nyhart (2016), "Today a more fitting image would be of the history of science as a densely tangled bank of people and material things teeming with social, cultural, economic, and religious life, that covers the globe." (p. 7). Researchers from the physics education field take up this movement without losing sight of the established orientation of the area for teaching NOS, taking HS as the guiding axis of analysis.

In this way, questions about science raised by studying material culture arose from SS are considered. For example, the Oldenburg group developed projects based on the reconstitution of the experiments, focusing on reconstructing the apparatus, recreating the experimental procedure and contextualizing the experience (Höttecke, 2000; and Heering, 2007). The replication process is used to study the techniques used to manufacture instruments and carry out experiments, the difficulties arising from the development of the experiment, and other issues inherent to the experimental construction. In PER, the replication process is used to develop NOS teaching considering the experimental aspects of science. Some projects are developed from physics teachers' training courses and others from the secondary-school level, as the ones developed at the University of Flensburg and the University of Campina Grande (Heering, 2015; and Silva *et al.* 2021). Considering the difficulty of having the original instruments in the schools and the potential of the science museums for teaching science, some scholars have developed projects for physics teaching in science museums (Heering and Müller, 2002; and Cavicchi, 2008).

Other researchers understand that the historical approach must address issues specific to the HS historiography field. In such a way, the central concern around historical approaches should not regard historical rigor or possible historical errors but the adherence to a historiographical perspective that contemplates questions intended to focus on physics teaching. In this perspective, studies expressing the Cultural History of Science (CHS) or Global History (GH) as the historiographical aspect is adopted in PER. As in Matthews (1994), it is assumed that science teaching loses meaning when HS is not evoked, but it is understood the imposition of new reflections on historical approaches to teaching due to the diversity of contemporary historiographical paths as well as issues specific to the field of science education.

CHS is referenced, for example, in Oliveira and Alvim (2021), when the authors present a proposal to teach NOS based on three dimensions: epistemological, sociocultural, and praxis. The authors propose to develop a sociocultural dimension based on CHS to discuss social and cultural issues which "have influenced (and do influence) the development of scientific and mathematical knowledge throughout history" (p. 744) and question the hegemony of science regarding other types of knowledge.

Historical approaches are advocated in teaching, aiming to broaden perspectives on the humanization of science. Based on studies carried out in schools in which students' histories were marked by colonialism and slavery, Moura and Guerra (2016), as well as Jardim, Guerra, and Fernandes (2021) argue that the CHS approach allows discussions about scientific practices and material and visual culture leading to an understanding of what has sustained science socially, culturally and materially. Not less important, CHS can point to who benefited and suffered in their formation. Thus, historical narratives with such bias allow discussion about science, emphasizing the participation of different social actors in it—such as women, artisans, illustrators, inhabitants of the colonies, etc.—making it possible to place science in the social, political, and historical spheres of the students' sociocultural context.

Focusing on GH, Sarukkai (2014); Gandolfi (2018, 2019); and Park and Song (2021) advocate a contextualization of science, and physics at a global stage, understanding scientific development as the result of global connections around the world. In this sense, the authors point to the importance of bringing historical narratives to science classes capable of showing that the scientific enterprise is not exclusive to Europe, just like it was not "born" in Europe and spread to other parts of the planet. Based on this, they highlight the importance of emphasizing the multicultural origins of modern science, especially the "transmission of knowledge from different traditions including the Chinese, Arabic, Indian, and the Greek" (Sarukkai, 2014, p. 1696).

Still, regarding the importance of emphasizing the multicultural origins of science, Gandolfi (2018) argues that the lack of diversity in educational proposals about cultures that were part of the construction of science ends up compromising one of the aspects defended by literature in the area as that derived from a historical approach to science: the humanization of science. For the author, the failure to explain that other cultures participated in the construction of modern science ends up presenting a partial humanization of the scientific enterprise, in which only white European men are those who participate in the scientific enterprise.

These movements of educators getting closer to the HS field seem to reflect questions raised by historians such as Chang (2021). He argues that different historiographical perspectives enable different political possibilities to understand science and its role in society. In this way, contemporary historiography is not understood as a cohesive body. The recent movement in the SE field that works at the HS and teaching interface suggests the existence of a diversity of possible historiographical affiliations for teaching. Each one implies different political perspectives for what is intended with science education. In this regard, HS is not presented only as a pedagogical tool that can be taken without the very field of HS historiography being evoked.

Movements in HS and philosophy of science fields, discussed in Sec. 9.2, indicate that just as understandings about aims and values in physics are related to questions and concerns present at a given time, the same can be said about values and aims for HS in PER and physics teaching. Dialoguing with views of aims and values in physics guided by the philosophy of science from the 20th century, in the 1990s, the focus of intended historical narratives for teaching was on theories considered fundamental for students to make decisions on scientific matters and how they developed and who were the natural philosophers and scientists who participated in the construction of these theories. In the 21st century, we recognize a movement in dialogue with the aims and values of physics brought by SS and new historiographical perspectives, pointing to problems of Eurocentric visions and considering possibilities of humanization of science broader than those initially taken by those working on HS and teaching frontier.

REFERENCES

Abd-El-Khalick, F. and Lederman, N., J. Res. Sci. Teach. **37**(10), 1057–1095 (2000).
Adúriz-Bravo, A. and Izquierdo-Aymerich, M., Sci. Edu. **18**, 1177–1192 (2009).
Allchin, D., *Teaching the Nature of Science—Perspective & Resources* (HSiPS Education Press, 2013).

Bachelard, G., *La Formation de L'Esprit Scientifique* (Vrin, 1993).

Bacon, F., *The New Organon*, edited by L. Jardine and M. Silverthorne (Cambridge University Press, 2003).

Bloor, D., *Knowledge and Social Imagery* (The University of Chicago Press, 1991).

Cavicchi, E . M., Sci. Educ. **17**, 717–749 (2008).

Chang, H., J. Gen. Philos. Sci. **52**, 97–114 (2021).

Clough, M., " Teaching the nature of science to secondary and post-secondary students: Questions rather than tenets," The Pantaneto Forum, Issue 25 (2007). See http://pantaneto.co.uk/teaching-the-nature-of-science-to-secondary-and-post-secondary-students-questions-rather-than-tenets-michael-clough/

Comte, A., *Cours de Philosophie Positive (Introduction et commentaires Florence Khoddos)*, PhiloSophie, Édition Électronique (Hatier, 2020). https://www.sudoc.fr/000674907.

Driver, R. *et al.*, *Young People's Images of Science* (Open University Press, 1996).

Daston, L. and Galison, P., *Objectivity* (Zone Books, 2007).

Erduran, S. and Dagher, Z., *Reconceptualizing the Nature of Science for Science Education Scientific Knowledge, Practices and Other Family Categories* (Springer Academic Publishers, 2014).

Fedi, L., *Comte (Figures du Savoir)* (Les Belles Lettres, 2017).

Fleck, L., *Genesis and Development of a Scientific Fact* (The University of Chicago Press, 1981).

Forato, T. C. M. *et al.*, Sci. Educ. **21**, 657–682 (2012).

Gandolfi, H., Sci. Educ. **27**, 259–297 (2018).

Gandolfi, H., Cult. Stud. Sci. Educ. **14**, 557–567 (2019).

Gay, P., *Weimar Culture: The Outsider as Insider* (W. W. Norton & Company, 2001).

Gil, D. *et al.*, Ciênc. Educ. **7**(20), 125–153 (2001).

Hacking, I., *Representing and Intervening: Introductory Topics in the Philosophy of Natural Sciences* (Cambridge University Press, 2010).

Hanson, N., *Patterns of Discovery: An Inquiry Into the Conceptual Foundations of Science* (Cambridge University Press, 1958).

Harding, S., *Objectivity and Diversity: Another Logic of Scientific Research* (The University of Chicago Press, 2014).

Heering, P. and Müller, F., Sci. Educ. **11**, 203–214 (2002).

Heering, P., Interchange **46**, 5–18 (2015).

Heering, P., Sci. Educ. **16**, 637–645 (2007).

Hobsbawm, E., *The Age of Revolution: 1789-1848* (Vintage, 1996).

Holton, G., Phys. Educ. **4**(1), 19–25 (1969).

Höttecke, D., Sci. Educ. **9**, 343–362 (2000).

Irzik, G. and Nola, R., Sci. Educ. **20**, 591–607 (2011).

Jardim, W. *et al.*, Sci. Educ. **30**, 609–638 (2021).

Kuhn, T., *The Structure of Scientific Revolutions*, 50th Anniversary ed (University of Chicago Press, 2012).

Lakatos, I. and Musgrave, A., *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science* (Cambridge University Press, 1970).

Lakatos, I., *The Methodology of Scientific Research Programmes*, Philosophical Papers Vol. I (Cambridge University Press, 1989).

Laudan, L., *Progress and Its Problems: Towards a Theory of Scientific Growth* (University of California Press, 1978).

Lederman, N. *et al.*, J. Res. Sci. Teach. **39**(6), 497–521 (2002).

Martins, A. F. P., Caderno Bras. Ensino Fís. **32**(3), 703–737 (2015).

Masterman, M., *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science*, edited by I. Lakatos and A. Musgrave (Cambridge University Press, 1970), pp. 59–89.

Matthews, M., *Science Teaching: The Role of History and Philosophy of Science*, (Routledge, 1994).

Matthews, M., *Advances in Nature of Science Research*, edited by M. S. Khine (Springer, Dordrecht, 2012), pp. 3–26.

Matthews, M., Sci. Educ. **1**, 1–9 (1992).

McComas, W. F. and Kampourakis, K., Rev. Sci. Math. ICT Educ. **9**(1), 47–76 (2015).

McComas, W. F. and Olson, J. K., *The Nature of Science in Science Education*, edited by W. F. McComas (Kluwer, Dordrecht, 1998), pp. 41–52.

McComas, W. F. *et al.*, *The Nature of Science in Science Education*, edited by W. F. McComas (Kluwer, Dordrecht, 1998), pp. 3–39.

McComas, W., Sci. Educ. **17**, 249–263 (2008).

Mill, J. S., *A System of Logic Ratiocinative and Inductive*, Collected Works (University of Toronto Press, 1981), Vol. VII.

Motterlini, M., *For and Against Method: Imre Lakatos and Paul Feyerabend* (Chicago University Press, Chicago, 1999).

Moulines, C. U., *O Desenvolvimento Moderno da Filosofia da Ciência (1890-2000)* (Scientiae Studia, 2020).

Moura, C. B. and Guerra, A., Rev. Bras. Pesqui. Educ. Ciênc. **16**(3), 749–771 (2016).

Nyhart, L., *A Companion to the History of Science*, edited by B. Lightman (Wiley-Blackwell, 2016), pp. 7–22.

Oliveira, Z. and Alvim, M., Caderno Bras. Ensino Fís. **38**(1), 742–774 (2021).

Osborne, J. *et al.*, J. Res. Sci. Teach. **40**(7), 692–720 (2003).

Park, W. and Song, J., Cult. Stud. Sci. Educ. **17**, 355–381 (2021).

Pestre, D., *Introduction aux Science Studies* (La Découverte, 2006).

Phillips, D., *A Companion to the History of Science*, edited by B. Lightman (Wiley-Blackwell, 2016), pp. 224–237.

Pickering, A., *The Mangle of Practices* (University of Chicago Press, 1995).

Popper, K., *The Logic of Scientific Discovery* (Routledge, 2002).

Reichenbach, H., *Experience and Prediction* (University of Chicago Press, 1938).
Rossi, P., *The Birth of Modern Science* (Wiley-Blackwell, 2001).
Sarukkai, S., *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. R. Matthews (Springer, Dordrecht, 2014), pp. 1691–1719.
Sevcenko, N., *A Corrida Para o Século XXI: No Loop da Montanha-Russa* (Cia das Letras, 2001).
Silva, A. P. B. *et al.*, Rev. Ciênc. Ideias **12**, 192 (2021).
Uebel, T., *Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, 2021). See https://plato.stanford.edu/entries/vienna-circle/.

CHAPTER

10 METHODS AND PRACTICES IN PHYSICS

Oliver Passon and Johannes Grebe-Ellis

Passon, O. and Grebe-Ellis, J., "Methods and practices in physics," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 10-1–10-26.

10.1 INTRODUCTION

In October 2017, the Court of Justice of the European Union (CJEU) ruled that the card game bridge is not a sport, since this would require "a not negligible physical element" (CJEU, 2017). However, the issue is controversial and for example, the World Bridge Federation is recognized by the International Olympic Committee (IOC).[1]

Why do we start with this apparently unrelated remark? What we have here is just another example for the infamous "demarcation problem," i.e., in order to write about "Methods and Practices in Physics," it would be good if one could define "physics" and distinguish it from other human activities and, especially, pseudo-science.

As with the question of whether bridge qualifies as a sport, drawing this dividing line turns out to be remarkably difficult. For some time (mainly in the 19th century) it was believed that a certain method of inquiry, the "scientific method" characterized the activity called "science" (or "physics"). However, as detailed below (Sec. 10.2), this is not considered seriously anymore. While there is certainly a set of characteristic activities (e.g., systematic observation, experimentation, inductive and deductive reasoning), there is apparently no fixed toolkit of methods which is common across science and only science (or, again, physics). It seems that in a situation like this a unifying account of science may either consist of overwhelming descriptive detail or trivial generalizations. A more optimistic take on this issue is given by Hepburn and Andersen (2021) in their Stanford Encyclopedia entry on "Scientific Method." They suggest that one may take a cue from the recent movement in philosophy of science

---

[1] This case was brought up by the English Bridge Union in order to obtain the tax reduction, which is granted to services linked to sport and physical education in the United Kingdom. I owe the bridge example to Joseph D. Martin (Durham).

https://doi.org/10.1063/9780735425514_010

toward a greater attention to practice, i.e., to what scientists actually do. As Hepburn and Andersen (2021) state:

> This "turn to practice" can be seen as the latest form of studies of methods in science, insofar as it represents an attempt at understanding scientific activity, but through accounts that are neither meant to be universal and unified nor singular and narrowly descriptive.[2]

So far, we have only sketched a debate within the philosophy (and history) of science; however, it parallels the development within physics education. Also here, there was (and still is) the tendency to claim a unique "scientific method" by which facts of nature are discovered.[3] Although it was acknowledged already in the 19th century that science provided no *certain* knowledge, scientific claims were still regarded as superior since the scientific method was viewed as sufficiently self-corrective. Or, as Ellery W. Davis put it in 1914: "If science lead (sic) us astray, more science will set us straight" (Davis, 1914, p. 49).[4]

Under this assumption, physics retains in essence its privileged epistemic status. This, in turn, implies that the study (and teaching) of the historical and contingent circumstances of the creation of scientific knowledge deserve only little attention. With this view, the organizing themes of physics teaching are consequently the "scientific method," the corresponding subject matter ("laws" and "concepts") and perhaps some anecdotal and glorifying historical accounts of the protagonists. In any event, with this view, the history of physics is of no relevance *per se* and at best a means to an end (see Sec. 10.4).

However, as indicated above (and explained in more detail in Sec. 10.2), there is broad agreement that such a privileged epistemic status cannot be justified strictly. Now, calls to make the "nature of science" (roughly speaking over and above the mere scientific facts) an integral part of science education can be viewed as a reaction to these exaggerated claims of the validity of science. Recognizing that physics is human-made (while mediated by the scientific community) and carries the traits of the knowing and perceiving subject, the place and time of its development as well as other contingent factors imply different or supplementary organizing themes for teaching physics. Importantly, all this does not imply that one has to submit to relativism or social constructivism, an extreme view from which even the originators have started to distance themselves [see, e.g., Latour (1992)].

"Methods and Practices in Physics" is still a vast subject and we cannot hope to do justice to this topic and need to be very selective. Still, in order to place the relevant issues into context, we start by a brief

---

[2] We note in passing that this shift to "practices" does not solve the demarcation problem. It is (also) only shifted to the question of what practice to consider. Or to put it differently, the slogan "Physics is what physicists do" assumes that we can tell the difference between a physicist and a non-physicist.

[3] The naïve "Scientific Method" may be construed as follows: (i) make observations, (ii) formulate a hypothesis, (iii) deduce consequences from the hypothesis, (iv) make observations to test the consequences, and, finally, (v) accept or reject the hypothesis based on the observations (Grandy and Duschl, 2007).

[4] Ellery W. Davis (1857–1918) was a mathematician at the University of Nebraska and made no important contributions to this debate himself. However, in the piece quoted, he reports on Charles S. Peirce's theory of induction, which tries to secure scientific knowledge by self-correction. The pragmatism of Peirce was certainly very influential at that time.

review of the demarcation problem (Sec. 10.2) since the issue of "the scientific method" arose in this debate. Subsequently, we turn to the nature of science (NOS) and scientific inquiry (SI) because an authentic and realistic image of physics research (i.e., its practices and methods) in physics teaching takes center stage (Sec. 10.3). The long tradition of the use (and abuse) of the history in teaching physics is briefly summarized in Sec. 10.4 before we turn to specific experimental and theoretical practices and methods in Sec. 10.5. The material there is largely structured by the discussion of common misrepresentations. In conclusion, we provide some reflections on the direction of future research in this field.

10.2 THE DEMARCATION PROBLEM AND THE RISE AND FALL OF THE "SCIENTIFIC METHOD"

In this section, we briefly go through some of the history of the demarcation problem as viewed through the lens of physics education. As noted by Gordin (2015), this problem can be viewed as "the central task of science pedagogy," since K–12 students should become literate enough to appreciate what it *means* to be scientific. In addition, the demarcation between science and non- or pseudo-science is not just an analytical but also a practical problem. As noted by Gieryn (1983), journal editors who reject some manuscripts as unscientific or education administrators who set up curricula which include chemistry but exclude alchemy routinely apply demarcation criteria.

An important text for the western tradition of science was the *Posterior Analytics* of Aristotle, in which he explained the criteria for *scientific knowledge*. According to Aristotle, the hallmark of scientific knowledge is its apodictic certainty, which follows from grasping the first principles and applying only logical demonstrations (Laudan, 1983). In addition, Aristotle demanded that scientific knowledge provide a causal demonstration, which marks it off from the crafts (which only know "how" and not "why").

These requirements for scientific knowledge changed significantly in the 17th century. While Galilei, Huygens and Newton still accepted the infallibility of science, they refused the need for causal demonstrations. Famously, Newton told his readers over and again that he did not engage in "hypothesis and speculations." His claim was that the infallible theories can be derived directly from these phenomena (Laudan, 1983, p. 114).[5]

Throughout the 17th and 18th centuries, European thinkers still regarded scientific knowledge as apodictically true and only the 19th century saw the emergence of the fallibilistic perspective in epistemology. Apparently also under the influence of actual sciences evolving, theories came to be

---

[5] A similar view among students was reported by Kang *et al.* (2004), who investigated Korean 6th–10th graders. Here, the majority of students (across all grade levels) viewed scientific theories as "facts which have been proven by many experiments." Given that this view is less often found among students from Western countries, the authors suggest an impact of cultural factors.

viewed as corrigible and subject to emendation. Hence, the stark contrast between “knowledge” and “opinion” became untenable. As a substitute for *certainty* as a demarcation tool, philosophers and scientists of the time suggested that the *methodology* distinguishes science from all other activities. But scientific claims were still regarded as superior since the scientific method was viewed as sufficiently self-corrective.[6] This position was, e.g., taken by Comte, Helmholtz, or Mach (among many others).

However, the project to identify a common repertoire of methods (let alone to establish the epistemic credentials of them) failed. Instead, there was a multitude of proposed methods (Laudan, 1981, Chaps. 8 and 10). But all these suggestions were not only divergent but also proved to be hard to explicate properly and, perhaps most devastating and relevant to our issue, they bore only little resemblance to the actual practices applied by working scientists, as noted already by Pierre Duhem in 1906 (Duhem, 1962). Laudan concluded that by 1900 neither certainty nor a privileged set of methodological rules could secure specific epistemic status of science beyond the rhetorical level. Ironically, the basis for privileged epistemic status was compromised exactly at the time when science developed rapidly and started to have a decisive impact on society.

Where do matters stand today? Laudan argued that also the attempts in the early 20th century to distinguish the scientific from the non-scientific (e.g., “verifiability” or “falsifiability”) failed. One of their problems lay in the fact that they neglected retrospective evidential assessment (Laudan, 1983, p. 122). Other candidates for the explication of this distinction do not fare better and Laudan argued compellingly that none of these suggestions provided necessary and sufficient conditions for the activity customarily called science. Laudan (1983) even suggested that the labels of “scientific” and “non-scientific” should be dropped—however, this is not to say that according to Laudan in each specific case claims of validity and reliability cannot (and should not) be debated. Presumably most philosophers (let alone scientists or science educators) disagree with Laudan’s suggestion to abandon the term “scientific” and, e.g., Pigliucci (2013) argued that one can learn from Wittgenstein how to deal with complex concepts that do not admit sharp boundaries. Wittgenstein’s notion of “family resemblance” is supposed to characterize exactly such cluster-concepts [compare also Erduran and Dagher (2014, Chap. 2) for applying the family-resemblance idea to conceptualize the nature of science within science education].

However, this view also admits that a grand unified methodology of science does not exist and that scientific claims are also, e.g., the result of processes of negotiation within the community (Hepburn and Andersen, 2021, Sec. 1). In conclusion, “scientific” is not just an epistemic category and “science” not just the sum total of objective facts which have been “discovered.” Science should rather be viewed as a particular epistemic and cultural practice, i.e., a complex activity embedded in historical, social and institutional contexts. All this implies that a sharp division between science and pseudo-science

---

[6] Again, this shows similarities with students’ conceptions about the nature of science. Tobin and McRobbie (1997) investigated the epistemic beliefs of Australian eleventh-graders. A majority took scientific knowledge to be subject to change and modification; however, they typically assumed that as a result of modification, a closer approximation to the truth is achieved.

cannot be made; however, this does not force one to submit to relativism or social constructivism. In fact, many subtle concepts cannot be defined rigorously while still being meaningful (the term "art" is a prime example) and the above-mentioned notion of "family resemblance" is one way to deal with this ambiguity. This recognition provides strong motivation to integrate the "nature of science" (roughly speaking over and above the mere scientific facts) as well as the "nature of scientific inquiry" into science teaching. In Sec. 10.3, we briefly summarize these approaches since they provide important background for our main concern. However, before doing so, the issue of "scientific method" deserves a second and closer look.

10.2.1 The "scientific method" as a rhetorical tool

While one may simply dismiss the idea of a "scientific method" (i.e., a fixed toolkit of methods which is common across science and only science) as a piece of misguided philosophy of science, it was pointed out by Thurs (2015) that it is much more interesting and revealing to view it as a *rhetorical strategy*.

He pointed to three related functions of the appeal to the "scientific method," namely, as a bridge to communicate science to lay persons, as a brand that represents science itself and as a tool for so-called "boundary work" in the sense of Gieryn (1983; 1999). In brief, this "boundary work" describes the efforts of scientists to create and shape the image of science to contrast it favorably to non-scientific activities. In a way, it represents the pragmatic aspect of the demarcation problem. The sociologist Thomas Gieryn (1983) investigated the mechanisms by which even apparently contradictory definitions of "science" have been exploited in order to enhance its access to social and material resources while denying such benefits to others. To this effect, the boundaries of science have been drawn (and redrawn) in flexible ways and the keyword "scientific method" has played an important role in this endeavor.

However, any term can play this keyword-function only if its definition is flexible enough to be a force of cohesion and inspiring action among groups (Thurs, 2015, p. 212). What is needed is just the right balance between precision and vagueness. Viewed this way, the vagueness (or even non-existence) of a "scientific method" proper was the key to its success as a rhetorical tool. Or, as Thurs (2015, p. 217) puts it: "Still, the scientific method did what keywords are supposed to do. It didn't reflect reality—it helped create it." In Sec. 10.4.1, we will again encounter the issue of boundary work in the context of distorted historical representations of physics.

10.3 NATURE OF SCIENCE AND SCIENTIFIC INQUIRY

A main goal of physics education is to support "scientific literacy," an elusive concept which may be roughly characterized as the ability to use scientific knowledge in order to make informed decisions in personal, social and political issues (Laugksch, 2000). However, in order to make such decisions, one needs more than scientific knowledge alone, but also knowledge about its characteristics (i.e., "being tentative while still highly reliable"), which in turn is related to the way it is generated (i.e., practices and methods of physics inquiry). In physics education, these issues are dealt with in the debate on the Nature of Science (NOS) and Nature of Scientific Inquiry (NOSI).

10.3.1 The consensus-based approach to nature of science and scientific inquiry

While it is generally acknowledged that the integration of NOS and NOSI into science teaching is desirable, its exact definition is debated. Clearly, philosophers, historians, or sociologists of science have no common view of the nature of science and inquiry, but there is certainly a core of established notions. Presumably, the dominant position in science education is the so-called domain-general and consensus-based view [see, e.g., Lederman *et al.* (2002) and McComas and Clough (2020)]. McComas and Olson (1998) found substantial agreement within a number of then recent science curriculum reform documents produced in the United States, United Kingdom, Canada, Australia, and New Zealand concerning the elements of NOS and NOSI which should be included in curricula. Also the Delphi study conducted by Osborne *et al.* (2003) could identify much common ground.

The corresponding NOS features are typically summarized in the form of lists. For example, according to Lederman (2007), one may state: scientific knowledge is tentative, empirically-based, subjective, necessarily involves human inference, imagination, and creativity, and is socially and culturally embedded. Furthermore, the specific distinction between observations and inferences as well as between scientific theories and laws is emphasized.

Based on this consensus-view, a NOS assessment instrument (Views of Nature of Science Questionnaire—VNOS for short) was developed (Lederman *et al.*, 2002). This instrument is widely used; hence, there are a number of empirical studies on the views of NOS students and teachers and how to teach NOS. Some important empirical findings of this line of research are that students (and teachers) typically do not possess "adequate" conceptions of NOS, that these conceptions are best learned through explicit, reflective instruction (as opposed to implicitly through experiences with simply "doing" science) and that teachers' valid conceptions of NOS are not automatically and necessarily translated into classroom practice (Lederman, 2007).

If one follows the consensus-based view, the NOS refers to the characteristics of scientific *knowledge*, i.e., a cognitive outcome. Consequently, it should not be conflated with scientific inquiry (NOSI). Lederman (2007, p. 835) suggested that the term "Nature of Scientific Knowledge" (NOSK) would have helped to keep NOS(K) and scientific inquiry distinct and Norman Lederman has continued to use this label [e.g., most recently, in Lederman and Lederman (2019)]. Now, Scientific Inquiry (or the "Nature of Scientific Inquiry" NOSI) refers to the *processes* of how scientific knowledge is developed (i.e., its source). As a student outcome, it may involve both doing inquiry (e.g., collecting data) but also knowledge about inquiry (i.e., planning an experiment). Another misunderstanding is to equate NOSI with inquiry as a mere *teaching method*, which puts emphasis on student activity, as often promoted by science education reform documents.

The Lederman group has also developed assessment-tools for scientific inquiry, namely, the *Views of Scientific Inquiry* (VOSI) (Schwartz *et al.*, 2008) and the extended version thereof, the *Views About Scientific Inquiry* (VASI) questionnaire (Lederman *et al.*, 2014). Within the VASI instrument, the construct NOSI has several aspects; some of them are (i) scientific investigations all begin with a question and do not necessarily test a hypothesis; (ii) inquiry procedures can influence results; and (iii) explanations are developed from a combination of collected data and what is already known. However, particular emphasis is given to the fact that there is no single and universal "scientific method" which can be followed algorithmically.

10.3.2 Alternative approaches to nature of science and scientific inquiry

The consensus-based approach promises an understanding of NOS and NOSI that is relevant and accessible to students. The proponents of this view stress further that the remaining disagreements about the precise definition or meaning of NOS and NOSI among philosophers, historians, and science educators are basically irrelevant to K-16 instruction (Lederman, 2007).

However, several authors have noted shortcomings and have emphasized the more controversial and context dependent character of NOS [see, e.g., Matthews (2012) or Duschl and Grandy (2013)]. Even Osborne *et al.* (2003), whose Delphi study found substantial agreement among their "expert community," expressed concern that the items might be regarded by teachers as *discrete* components, taught in an abstract way only. Also, the consensus-view neglects, e.g., differences among scientific disciplines and portrays NOS as fixed and timeless (Irzik and Nola, 2011).

One specific criticism targets NOS as an educational goal at all grade levels. Elby and Hammer (2001) argue that one should distinguish between the *correctness* and the *productivity* of an epistemological belief. While the sophisticated view that scientific results are tentative (rather than objective and fixed) is accepted in the philosophy of science, it may pose a disadvantage in the learning process. For example, a student of Newton's axioms would be ill-advised to blame their counterintuitive implications on their only "tentative" nature.

In a similar vein, Douglas Allchin has pointed out that the consensus-view of NOS may run into the danger of focusing too narrowly on declarative knowledge (Allchin, 2011). He also sees elements missing, e.g., funding, motivations, peer review, cognitive biases, and fraud. A central role is played by historical (or recent) case-studies which are supposed to foster a contextualized understanding of the nature of science for the students. In this sense, Allchin's approach is more inductive in spirit compared to the consensus-based approach, which starts out from the more abstract items on the NOS-list.

In light of this controversy, Irzik and Nola (2011) have provided still another conceptualization of NOS. Here, the notion is based on the concept of family resemblance; see also Erduran and Dagher (2014). This approach is much broader than the consensus view, and (according to Lederman) it encompasses the aspects identified in the consensus lists (Priemer and Lederman, 2021, p. 130).

It also proved to be instructive to investigate what scientists themselves say about NOS and scientific inquiry—as the interview study by Wong and Hodson (2009) did. In this study, 13 well-established scientists from various fields completed a modified version of the open-ended questionnaire, *the Views of Nature of Science Questionnaire* (VNOS-C), developed by Lederman *et al.* (2002). Also, here, some tension to the consensus-based view arose (Wong and Hodson, 2009, p. 123). In particular, they find fault with the categorization of NOS views as either "naive" or "adequate" on the basis of a questionnaire response.

As we have discussed in Sec. 10.3.1, the consensus-view stresses the danger of conflation of NOS and NOSI. Roughly speaking, all of the alternative conceptualizations of NOS criticize this strict separation between NOS and NOSI as artificial [e.g., Irzik and Nola (2011)]. Or as Allchin (2012, p. 696) puts it: "In science, however, process and product are virtually inseparable."

The reader has to keep in mind that the results based on assessments with the VNOS, VOSI, or VASI instruments are tied to their specific way to spell out the nature of science or scientific inquiry and cannot be easily generalized to different conceptualizations. However, this controversial debate should not obscure the fact that the different conceptualizations of NOS and scientific inquiry also share many common features. For example, all of them acknowledge that physics teaching needs to be informed by the history of physics. This approach can certainly build on the venerable tradition.

10.4 THE HISTORY OF PHYSICS IN PHYSICS TEACHING

Even many years before "scientific literacy," "nature of science" or "scientific inquiry" became catch phrases in science education, physicists tried to improve their teaching by including material from the history of physics into the curriculum. This line of thought can be easily traced back to thinkers like Pierre Duhem (1861–1916) or Ernst Mach (1838–1916) who supported a historical method (or genetic approach) in teaching. Many of the more recent developments can be traced back to the pioneering work of James B. Conant since the late 1940s (Conant and Nash, 1957). He introduced a historical case-study approach into the college courses for non-science majors. Originally developed for Harvard, this approach was widely adopted for other non-science major programs in the U.S. (Matthews, 1992, p. 15).

However, for physics major programs, these developments were largely inconsequential (Matthews, 1992). Conant's student (and later fellow Harvard Professor) Leo Klopfer adapted the case study approach to high schools (Klopfer, 1964); however, this had no lasting impact in the USA or abroad. High school curricula of the 1960s were largely science content oriented and neither the history nor the philosophy of science was considered much (Duschl, 2006).

Höttecke and Silva (2011) have pointed out major obstacles for the successful implementation of history and philosophy of science (HPS) elements into the classroom. One of them is the lack of *adequate* HPS content in textbooks. The situation is even worse than that, since there is a long and still prevalent tradition of including *inadequate* (i.e., anecdotal and distorted) history into physics textbooks. Here, the use of history is potentially harmful to the understanding of the practices and methods of physics.

10.4.1 The misuse of history in physics teaching: Historical framing and presentism

There is a long tradition of just *framing* science teaching "historically" by including anecdotal reference to historical events or by presenting a streamlined account which is reading present beliefs back into the past (Franklin, 2016). As Klein (Brush and King, 1972, p. 13) complained, these superficial accounts distort the actual course of events, ignore the gradual conceptual changes and present the history as a cumulative sequence, which finally led to the acceptance of the current theories. In short, this "Whig-history" [also called "quasi-history" (Whitaker, 1979) or historical "myth-conceptions" (Allchin, 2003)] fails to provide an appropriate picture of the nature of science and the practices and methods employed. The wide prevalence of these narratives raises the question of how and why they originated.

Klein (Brush and King, 1972) argued that the origin of distorted historical accounts in physics teaching had structural reasons since mostly those aspects of the history that are relevant to the current curriculum were included. Clearly, this ignores what "[…] really concerned past physicists, the context within which they worked, or the argument that did or did not convince their contemporaries to accept new ideas" (Brush and King, 1972, p. 13). This claim has important implications for any use of history in physics teaching. As long as the subject matter functions as the controlling element, the danger of "presentism" as described by Klein is real. At the same time, it needs to be acknowledged that this is rather a debate about avoiding *severe* distortions and including *more* contextualization since some sort of presentism, Whiggism (Hall, 1983) or anachronism (Brush, 2021) is certainly inevitable and may even be desirable.

In Sec. 10.2.1, we introduced the notion of the "scientific method" as a rhetorical tool. Something similar applies to quasi-history, which, by the way, often contains an idealized image of the scientific method likewise. Already in 1974, the historian of physics Stephen Brush had suggested that the distortion of the historical events is also motivated by the wish to convey a specific image of the "correct attitude or general methodology" (Brush, 1974, p. 1164) of science. Brush argued that an honest account of the history makes it hard to convey the myth of scientists as "neutral fact finders" and science as an objective enterprise, driven by experimental evidence and logical rigor only. Brush is referencing especially the works of Thomas S. Kuhn, who has commented explicitly on the functional role of historical accounts in science textbooks. Kuhn wrote: "Characteristically, textbooks of science contain just a bit of history […]. From such references both students and professionals come to feel like participants in a long-standing historical tradition. Yet the textbook-derived tradition […] never existed" (Kuhn, 1996, p. 137f).

In a recent book José G. Perillán (2021) has lifted the debate about the relationship between the history of science and myth-historical representations to another level. He demonstrated that many physicists deliberately apply distorted historical narratives. For example, Richard Feynman confessed in a popular science book: "By the way, what I have just outlined is what I call a 'physicist's history of physics,' which is never correct. What I am telling you is a sort of conventionalized myth-story that the physicists tell to their students, and those students tell to their students, and is not necessarily related to the actual historical development, which I do not really know!" (Feynman, 2006, p. 6). A similar remark can be found in the popular science book *The God Particle*, authored by the American physicist Leon M. Lederman (not to be confused with the above mentioned Norman G. Lederman). Although hidden in the postscript on page 412, Lederman confessed openly: "However; from the point of view of storytelling, myth-history has the great virtue of filtering out the noise of real life. [...] There may, in fact, be no source for some of the best stories in science, but they have become such a part of the collective consciousness of scientists that they are 'true,' whether or not they ever happened" (Lederman and Teresi, 2006, p. 412).

In describing the function of the myth-history as "filtering out the noise of real life" Lederman[7] is apparently subscribing to the view that physics is a self-correcting enterprise which will inevitably approach the truth (i.e., the position of the 19th century; compare Sec. 10.2). This, however, implies that the benefit of a faithful historical presentation is very limited since the contingent historical facts do not alter the scientific results which were (and will be) eventually discovered. Or, as Feynman is quoted from a TV documentary from 1973 (Perillán, 2021, p. 42): "The real test in physics is experiment, and history is fundamentally irrelevant."[8]

The aforementioned quotes show that a mythical and idealized image of physics is not only the result of accidental slips by thoughtless textbook authors but is actively promoted by some of the most distinguished physicists. Our own material in Sec. 10.5 on experimental and theoretical practices is organized in a large part by myth-historical distortions which need to be corrected. However, Perillán (2021) makes a compelling case that historians (or physics educators for that matter) should do more than merely *correct* these myth-histories—they should *study* them since they originate for telling reasons. Eventually, "they help create scientific consensus by amplifying a preferred historical signal. They filter out and further marginalize people and ideas that donot align with the status quo" (*ibid.*, p. 69). That is, they have not only a huge impact on public discourse, physics teaching, and directions of future research but also reveal the internal power structure of science.

At the same time, Perillán reminds us again that *all* forms of historical inquiry are subject to presentism or anachronism to a greater or lesser extent (Hall, 1983; and Brush, 2021). Therefore, Perillán (2021, p. 70) suggests that the relationship between "history" and "myth" should not be reduced to a simple polarizing conflict. This is closely related to the following issue. Even a properly contextualized case study (which virtually avoids myth-historical elements completely) may be non-representative of the practices and methods of physics. For example, many debates in the philosophy of science argue with a

---

[7] Lederman was an experimental high-energy physicist, i.e., improving the signal-to-noise ratio apparently became the second nature to him.

[8] This TV documentary on Feynman is called "Take the World from Another Point of View" and can be accessed via YouTube. The remark is made after eight minutes.

specific sample of historical case studies. Larry Laudan's "pessimistic meta-induction," e.g., is supposed to show that we have no reason to have confidence in the (approximate) truth of our current theories because there are many historical counterexamples of refuted theories once believed to be true. As argued convincingly by Mizrahi (2015), this and similar arguments suffer from a selection bias in their sample of case studies. Bolinska and Martin (2021) note (also in the context of history and philosophy of science) that the use of a specific "canon" of case studies is only justified on the assumption that these are "more representative, more influential, or otherwise more important than the vast majority it neglects." These authors suggest principles to guide more effective canonization practices, i.e., to avoid what they call the "tragedy of the canon." A similar debate on the *selection criteria* of case-studies and vignettes needs to be held within the community of historically oriented physics educators. But physics education may face an additional problem here. There are good (educational) reasons to select case studies *because* they are non-representative, i.e., for their extraordinary, memorable, or startling character. However, this is a pervasive problem (if one wants to call it a problem at all) not limited to science. Also, English classes prefer, say, Doris Lessing to a more "representative" (i.e., mediocre) author.

10.4.2 The use of history in physics teaching: recent developments

Even if physics textbooks at the high school and college level would avoid (or—given what we have said before—if their authors would be willing to avoid) the quasi-historical distortions discussed above, this would not translate into a more effective teaching of the subject matter or the NOS on its own. What is needed is a conceptual framework and empirical evidence on how to implement the history of physics into curricula and classroom activities. We can be brief here because these issues are dealt with in Chap. 9 by Elizabeth Cavicchi.

Seroglou and Koumaras (2001) and McComas (2011) provide a typology and comparative presentation of the various proposals concerning the contribution of the history of physics in physics education. Substantial work has been done on implementing historical vignettes and case studies as contexts in physics teaching; see Stinner *et al.* (2003) and Metz *et al.* (2007). The corresponding activities range from story-telling and creative writing to developing stage plays.

Höttecke *et al.* (2012) report on the results of the European Union funded HIPST project (History and Philosophy in Science Teaching, 2008–2010). Here, the focus was on the development of teaching and learning material, while science teachers were systematically integrated into the developmental work to enhance their attitudes and professional skills. A somewhat related approach is developed rather independently by a small but active community in Germany and Switzerland under the heading of "Lehrkunstdidaktik" (roughly translatable as "Art of Teaching Approach"). This group was founded by Wagenschein student Hans Christoph Berg and has produced various teaching units (called "Lehrstücke"). These teaching units—typically based on historical events—are carefully documented and downright performed almost like a stage play—in which the students participate actively [see, Berg (2004) and Emden and Gerwig (2020)].

10.5 PHYSICS AND PHYSICS EDUCATION ON PRACTICES AND METHODS

It is rather uncontroversial to subdivide the practices and methods of physics into experimental and theoretical activities. Historically, this distinction was not always sharp and some fields still operated at the boundary.[9] However, for roughly one century, this distinction has been institutionally anchored, e.g., in the designation or denomination of institutes and chairs or by specialized journals.

In physics education at the high school and college level, the distinction between experimental and theoretical physics is typically not made explicit. At first glance, this may seem like a problem for giving an appropriate and authentic picture of physics research [e.g., in high energy physics; see Wong and Hodson (2009, p. 117)]. However, one should bear in mind that for centuries this division of labor did not exist. Thus, to emphasize the unity of physics and the close interplay between experimental and non-experimental practices makes a lot of sense. In fact, the recent scholarship, especially on experimentation and the role of instruments, has made it abundantly clear that any sharp distinction between experimental and theoretical practices is problematic anyway (see Sec. 10.5.1).

Having said this, we still choose the categories "experimental" and "theoretical practices and methods" in order to structure our presentation. But we should mention another caveat from the outset. When the request to "learn physics" by "doing physics" is applied to experimental practices, the tacit assumption seems to be that the classroom experiment resembles the experiment in actual research. In specific open-ended inquiry-based learning settings this may even be true, but most experimental activities in the classroom remain largely pre-structured. Chinn and Malhotra (2002) conducted a comprehensive study on the deviation between methods in science classes compared to scientific laboratories. They conclude "that much work to be done to transform schools into places that nurture epistemologically authentic scientific inquiry" (*ibid.*, p. 214). See also Höttecke (2013) for a careful examination of this problem.[10] While we keep this distinction in mind, we turn to the experimental practices and methods.

10.5.1 Experimental practices and methods

As, e.g., the draw-a-scientist test reveals (Chambers, 1983), the laboratory and experimental equipment are the stereotypical workplace and identifying feature for scientists. But not only, this stereotype neglects the various other working environments of scientists. Additionally, textbooks, classroom practice, or popular news outlets often misrepresent experiments and experimenters. In turn, misconceptions about the nature of scientific inquiry and practices are promulgated.

---

[9] For example, astronomy cannot perform experiments and the counterpart of the theoretical activity is "observation." In addition, the area of "computational physics" has gained a rather independent position between experimental and theoretical physics.

[10] One should not, however, ignore the fact that classroom experiments (or "laboratory tasks") may also differ deliberately from research experiments.

The historian of physics Klaus Hentschel (2003) lists common "myths" with respect to experiments. Among them are (i) the apparent simplicity of carrying them out (especially, reinforced by elegant demonstration experiments—a development which can be traced back to the Age of Enlightenment and continues to classroom demonstrations today), (ii) their presentation as merely testing a theoretical hypothesis, or (iii) their alleged ability to distinguish between competing theories (experimentum crucis). He further criticizes (iv) the overemphasis of the role of accidental discoveries (serendipity) and (v) the common neglect of assistants and laboratory technicians.

Initially, the philosophy of physics in the early 20th century was concerned foremost with the development of theories, e.g., the then recent relativity and quantum physics. This tradition did not ignore experiments completely, but trivialized them largely as the mere reading of a pointer position. This is particularly ironic, given that, e.g., the logical empiricists and their followers devoted much of their attention to the distinction between observables and unobservables, and the epistemic bearing of observational evidence on theories (Boyd and Bogen, 2021).

Hentschel's critique reflects the more recent (i.e., since the 1980s) scholarship in the history and philosophy of science (HPS), i.e., the "new experimentalism" with its greater emphasis on material objects such as scientific instruments (e.g., Franklin, 1989). In what follows, we draw on this scholarship and use Hentschel's list of misrepresentations of experimental practices to structure our presentation. Where necessary, we will discuss related issues as well.

10.5.1.1 The apparent simplicity of experiments

Present demonstrations can easily obscure the (conceptual and technical) difficulties encountered when the experiments were conducted for the first time. This holds, in particular, for historical experiments with modern technologies unavailable at that time. For example, one may think of a modern free-fall experiment with a light barrier in contrast to Galilei's inclined plane experiment or of an 8th grader easily verifying Ohm's law with the help of a digital multimeter, which assumes the validity of this law for its functioning from the outset. As a result, the findings may seem trivial and the activity fails to provide an authentic view of the scientific practice.[11]

One way to avoid this misrepresentation is to reenact the experiment with a reconstructed device. The role of this replication method, championed by the Oldenburg group in Germany, in particular, established by Falk Rieß, is of course more far reaching (see, Heering and Höttecke, 2014). This method for authentically reconstructing material procedures and instrumental manipulations can be used in

---

[11] Interestingly, when it comes to current research, the situation can also be reversed. In most areas of current science, computer technology for data logging and analysis has made experiments less cumbersome than suggested using the "old school" methods applied in classrooms or lab classes. These innovations can have a huge impact on the experimental and theoretical practices. In their interview study, Wong and Hodson (2009) reported that the materials scientist, medical geneticist, and virologist all commented that the much reduced data generation time makes meticulous hypothesizing and theorizing before actual experimenting less important than in the past.

order to scrutinize their general relationship to the development of theories or models and to reveal tacit knowledge and skills not contained in the written documents. Thus, the reconstruction method can genuinely contribute to understanding the history of physics. For example, Peter Heering investigated Coulomb's inverse square law using the reconstruction method (Heering, 1992). Curiously (to the modern reader), Charles Augustin de Coulomb (1736–1806) had only published three measured values as "empirical proof" for his $\frac{1}{r^2}$-law and the replication of the Oldenburg-group showed that the torsion balance replicated according to source information creates considerable metrological problems. Apparently, Coulomb's design of the experiment and the data analysis had already presupposed the validity of the inverse square law while textbooks disseminate the myth of this law being based on solid empirical evidence found by Coulomb.

Other specific case studies have been conducted (Heering, 1994; 2005). Heering (2000) reported the positive experience of teaching electrostatics in secondary school based on replicated experiments; see also Chap. 9 of this handbook by Elizabeth Cavicchi.

While all of the above examples (and most of the work done in this field) concern "classic experiments" of the past, the replication method may be extended to lesser known or even forgotten experiments as well. Only recently Hasok Chang has made a compelling case that these forgotten experiments from the past can even recover lost scientific knowledge and extend what has been recovered. Hence, Chang challenges the view that scientific progress is strictly cumulative and has coined the term "complementary science" for this realm of lost knowledge. In Chang (2011), some fascinating and elementary examples (e.g., the boiling point of water in so-called super heating) are discussed and a connection to science education is drawn. Chang claims that in dealing with these anomalous results, students have "a genuine experience of inquiry and take a live lesson in NOS from that experience" (*ibid*., p. 335). Of a similar kind is the work of Rang and Grebe-Ellis (2018) on inverse spectra, which is also inspired by some "lost science," namely, Goethe's theory of color.

10.5.1.2 The experiment as mere hypothesis testing

An important NOS issue concerns the specific role of experiments in the generation and confirmation of scientific knowledge. While in the 19th century an inductivist view was still advocated by some, the work of Duhem (1962) from 1906 made it clear that such a view is actually untenable.[12] Prominent non-inductivist views on *confirmation* in the 20th century were the hypothetico-deductive (H-D) method (not to be confused with the deductive-nomological model of *explanation*) and falsificationism.

Falsification is deductive and similar to the H-D method in that it involves scientists deducing observational consequences from the hypothesis under test. However, as the name suggests, it stresses the falsification and not the confirmation (Hepburn and Andersen, 2021; Secs. 10.3.2 and 10.3.3). But in both cases (and this is the key issue here) the relation between "theory" and "experiment" is one-sided.

---

[12] However, Matthews (1989, p. 4) criticized that such an "inductivist model of inquiry" is still prevalent in many classroom activities.

An influential treatment of this position was given by Karl Popper in 1934. In his *The Logic of Scientific Discovery*, only a few pages were devoted to the relation between experiment and theory and Popper claimed that the sole function of experiments is to answer the specific questions asked by the theory, i.e., testing (and possibly falsifying) hypotheses (Popper, 2002, p. 89f). As a perfect example, Popper quoted the confirmation of de Broglie's matter-wave hypothesis by Davisson and Germer in 1927.[13]

This notion can be criticized on several grounds. For one thing, it assumes a strict separation between experiment and theory, which is questionable given the issue of theory-ladenness as championed by Norwood R. Hanson (1924–1967). The idea here is that theoretical presupposition can affect the observation on various levels (say, expectations, functioning of the instrument, etc.). For example, early "thermometers" were not sealed, i.e., the reading was affected (or even dominated) by the surrounding air pressure. A reliable "experimental" measurement of the temperature could only be gained after a sufficient theoretical understanding (Middleton, 1966, p. 28). This clearly undermines the categorical distinction on which the idea of hypothesis-testing is based (Boyd and Bogen, 2021).

But even if one accepts the theory-experiment distinction on pragmatic grounds and acknowledges that experiments as hypothesis-testing occur (e.g., when in 1819 Dominique Arago tested Poisson's prediction of a bright spot at the center of a circular object's shadow), it can hardly be the *only* function of an experiment. For example, the calibration of an instrument, the experimental determination of a numerical parameter, or the use of theories as a heuristic tool within the search for a new effect would be of a different kind. However, all the aforementioned types of experiments are still driven by specific theories. This leaves completely open how experimental practices can contribute to a field in which no theory or even a conceptual framework is available.

Case studies of the historian Friedrich Steinle (1997, 2006) have brought him to introduce the new category of "exploratory experiments" in contrast to the then "standard view" of "theory-driven experiments." In the latter case, experiments are done with a well-formed theory in mind, which affects their specific design, their execution and finally their evaluation – all potential sources of theory-ladenness by the way. In contrast, exploratory experimentation is driven by the desire to obtain empirical "regularities" (not yet theories) and to find out concepts and classifications by means of which those regularities can be formulated in the first place. According to Steinle (1997), exploratory experimentation typically occurs in those periods where research fields are still uncharted.

While according to Steinle, exploratory experimentation should not be viewed as a specific and well-defined procedure, one may nevertheless identify some common features. Among them is the variation of experimental parameters, the determination of which of the different experimental conditions are indispensable and the search for stable empirical rules (and of concepts by which these rules can

---

[13] This happens to be a poor example, given that de Broglie did not publish any suggested experiment in order to test his hypothesis. Furthermore the experimental study by Davisson and Germer was conducted completely independent of de Broglie's conjecture (Gehrenbeck, 1978).

be formulated). Most importantly, Steinle claimed that this type of experimentation has epistemic significance. The "standard view" may underestimate this significance by claiming that these case studies explore only the "context of discovery," while the epistemic value of experiments lies in the "context of justification." However, exploratory experimentation leads to the forming and stabilizing of conceptual frameworks, i.e., provides the very language in which the theories subsequently are formulated (Steinle, 1997, p. S72).

While exploratory experimentation may often function as a precursor for more theory-driven experiments, this practice has also been noted in other contexts. Steinle has already observed that, e.g., Goethe's theory of color is based on a similar approach (Ribe and Steinle, 2002).[14] Goethe's ambition, however, was not directed towards a theoretical description in the modern understanding and his approach is often dubbed "phenomenological." However, this approach to nature has been exploited in physics education. Famously, Rudolf Steiner, the founder of Waldorf education, was strongly influenced by Goethe's ideas on nature (Steiner, 1988). Dahlin (2001) provides additional underpinning for this approach and Østergaard *et al.* (2008) review the phenomenological work in science education. A recent study by Park and Song (2018) refers to Goethe's concept of the "experiment as mediator" as fostering practical activities in physics classrooms. They stress that it is "interconnecting individual observations by means of repeating an experiment with varied conditions and levels of simplification and emphasizing the close contact and interaction of the observer to the object of inquiry." These, however, are exactly some of the typical steps in an exploratory investigation—in contrast to the often criticized "cookbook-style implementation, piecemeal empiricism, and students' alienation from nature" (*ibid*., p. 57).

A research project by the Pedagogical Research Centre at the Federation of Waldorf Schools in Kassel (Germany) and the Alanus University of Arts and Social Sciences in Alfter (Germany) has given rise to a number of publications of fully worked out teaching units in mechanics, thermodynamics and electricity for various grade levels (Sommer 2020, 2021, 2022). All these phenomenological approaches possess a degree of similarity to inquiry-based science education (IBSE) and a recent large-scale assessment relates the high motivation of Waldorf students to this feature (Salchegger *et al.*, 2021). However, at the same time, only moderate science achievement is observed among the Waldorf students and the authors suggest that this may be due to an ineffective IBSE teaching strategy (*ibid*., p. 17).

This brings us to the question whether this exploratory and inquiry-based teaching which is in some respect closer to actual research practice, is effective in terms of science achievement and attitudes. Several studies have shown that a moderate level of inquiry teaching is more efficient than a high level of inquiry teaching regarding students' science achievement (Jiang and McComas, 2015; and Teig *et al.*, 2018). These findings already indicate the restricted efficiency of inquiry-based teaching. Only recently,

---

[14] As indicated at the end of Sec. 10.5.1.1, Goethe's theory of color can also be viewed as "complementary science" in the sense of Hasok Chang. For a recent discussion on the relation between Goethe's theory of color and modern optics, see Grebe-Ellis and Passon (2020).

Zhang *et al.* (2021) even went so far as claiming an "evidence crisis" in science educational policy. They argue that almost all *controlled* studies have found only minimal support for teaching science through exploration-based investigations. To the contrary, these studies support the recommendation that various forms of explicit instruction are more effective in terms of science achievement. Zhang *et al.* (2021) acknowledge the importance of other learning outcomes (i.e., attitude or interest in science), which may be supported by inquiry-based methods more effectively.

10.5.1.3 The "experimentum crucis"

The idea that single experiments can bring about a *final* decision between competing theories (and lead to the acceptance of the successful theory) can be traced back to Francis Bacon (1561–1626). It can be viewed as an extreme type of a theory-driven experiment and suffers from similar problems. Especially, the already mentioned issue of theory-ladenness casts doubts on this alleged function. A related issue is the underdetermination of theory by experiment (or "data"). As noted by (Quine, 1953, p. 41): "[…] our statements about the external world face the tribunal of sense experience not individually but only as a corporate body." Now, according to the Duhem–Quine thesis (i.e., glossing over important differences between these authors), any hypothesis assumes additional background assumptions. Hence, if a specific experiment fails to confirm the prediction, one may blame (and abandon) one of these auxiliary hypotheses instead. On this view, confirmation is a holistic enterprise and the experimentum crucis impossible.

Still, physics textbooks are full of experiments which allegedly provided "undeniable evidence" for specific theories or theoretical entities. For example, the photoelectric effect or Compton-scattering are frequently quoted as unambiguous confirmations of the light-quantum hypothesis. These specific examples ignore relevant auxiliary hypotheses. Both effects can be accounted for in the semi-classical theory, i.e., by treating only the matter quantum mechanically while the electromagnetic field is described classically (Passon and Grebe-Ellis, 2017).

How then do theories become accepted knowledge? The historians Brush and Segal (2015) used comprehensive case-studies to investigate exactly this question and revealed numerous relevant factors. The main focus was on the specific aspect of whether a theory was accepted because (i) it leads to a successful *prediction* or (ii) it provides a successful *explanation* of facts already known but not understood. Their case-studies showed that both of these instances occur—and sometimes a theory was even accepted without specific empirical support [e.g., special relativity, see (Brush and Segal, 2015, p. 82)] In the case of quantum mechanics (QM), the acceptance was apparently not based on new predictions (some of them where corroborated only years later) but on the successful explanation of known facts. The only one who seemingly took offence to that was the Nobel Prize committee. Brush and Segal suggest that the belated awards to its founders were due to missing predictions-in-advance. This mistake was corrected "only" in 1932 (Heisenberg) and 1933 (Schrödinger)—but not due to successful QM predictions in the meantime, but because of the discovery of the positron, i.e., a successful prediction of Dirac's relativistic version of QM (Brush and Segal, 2015, p. 495).

10.5.1.4 Serendipity

While accidental discoveries in physics have happened (e.g., Galvani's discovery of the electric properties of frog legs), this is certainly not the norm. However, especially among young students [e.g., grade 7 as observed by Meyer and Carlisle (1996)], the notion of experimentation as merely "trying out" in order to discover something new is rather common. Gyllenpalm and Wickman (2011) analyzed the understanding of experimentation among Swedish pre-service teachers and found that their views were not much more sophisticated either. Asked "about the meaning of 'experiment,' this term was given everyday connotations like 'trying' or 'testing' in a nontechnical sense" (*ibid*., p. 920). In addition, the pre-service teachers did not distinguish between experimentation as a scientific activity in contrast to a "laboratory task" as an educational activity. This view clearly neglects, e.g., how experimental practices are embedded in research programs and performed under controlled conditions. However, if the dominant view takes experiments as theory-driven (compare Sec. 10.5.1.2), the only way in which truly new experimental discoveries can be made is by accident. In this way, the standard-view of experiments as mere hypothesis-testing may reinforce this common misconception.

10.5.1.5 The neglected role of assistants and laboratory technicians

In 1935, Bert Brecht wrote the poem "Questions of a reading worker." It contains the verses: "Young Alexander conquered India./He on his own?/Caesar defeated the Gauls./Didn't he at least has a cook with him?" What Brecht is alluding to is the well-known phenomenon of "great man history," which neglects not only "great women" but also lesser known individuals. Something similar holds for the myth-history of physics with its strong emphasis on "great man," thus largely ignoring the contributions of women but also technicians and assistants (regardless of gender). When in the same context Steven Shapin (1989, p. 556) asked "Who did Boyle's experiments," he almost seemed to add a line to Brecht's poem.

Some of Boyle's technicians or laboratory aids may have just carried out unskilled physical work, which does not justify any specific acknowledgment. However, in at least one case the assistant (named Denis Papin) was involved in the design, performance and even documentation of experimental work while receiving no due credit (*ibid*., p. 559). These cases mark the two poles of a continuum in which Shapin brings up issues such as "moral economy," class distinctions and tacit skills. Such skill was clearly involved since technicians were often pointed to (i.e., they became visible) when "matters did not proceed as expected" (*ibid*., p. 558). Experimental failures were often attributed to defective instruments or the mishandling by assistants, while successful observations were simply treated as "visible testimony of nature" (*ibid*., p. 558).

In that sense, the neglect of these instrument makers, technicians and assistants leads to a "distorted and impoverished understanding of scientific practice" (Shapin, 1989, p. 563). Importantly, this holds not only for the past history but also concerns the understanding of scientific practices today. In Sec. 10.5.2.2, we will come back to a similar issue in the context of theoretical (or general) practices in physics.

10.5.2 Theoretical practices and methods

The request to "learn physics" by "doing physics" is also not easy to comply with when it comes to the theoretical practices. While some of them (e.g., problem-solving, generating hypothesis or providing explanations based on known theories) are within the reach of students, the *development* and *formulation* of physical theories is hard to retrace by student activities.

Scientific theories are typically not only presented in their "final form" (Duschl, 1990, p. 68), but the anecdotal history of physics is full of romanticized episodes where theories are apparently born fully fledged in the mind of a scientific genius. Famous examples are Archimedes' shout of "Eureka" from his bath, Newton's falling apple giving rise to his theory of gravity, or the youthful Einstein chasing in his imagination a beam of light and thereby anticipating the theory of special relativity. We will have a closer look at these "discovery myths" (Sec. 10.5.2.1) and the notorious "great man history" (Sec. 10.5.2.2) before we turn to the debate on scientific models and modeling (Sec. 10.5.2.3)—an important activity in physics teaching which makes close contact with theoretical practices and methods.

10.5.2.1 Discovery myths: Newton and Einstein

With respect to Newton's apple tree story, Fara (2015) notes that "The factual truth […] is not particularly important: what matters is its symbolic significance as the founding moment of Newtonian physics" (*ibid.*, p. 51). In particular, the biblical allusion suggests the identification of Newton with a "new Adam" and the sudden insight beneath the apple tree is the "intellectual equivalent of a divine visitation" (*ibid.*, p. 51). Famously, the British poet Alexander Pope (1688–1744) wrote on the occasion of Newton's death the lines "Nature and Nature's Laws lay hid in Night. *God said,* Let Newton be! *and All was* Light" [quoted from Fara (2015)]. Echoing the biblical account of the creation is again stylizing Newton as a Christlike figure.

The apple-episode became popular in the 19th century, when science and religion were increasingly in opposition to each other. Fara (2015, p. 55) suggests that this popularity also illustrates how (scientific) "genius took over the cultural significance and functions formerly attributed to sanctity." While there is no evidence that the falling apple played any role in Newton's discovery whatsoever, this myth (and similar accounts of sudden inspiration) still holds its grip. It is hard to imagine how under this influence "ordinary" students can identify with such a scientific genius and gain a sober understanding of theoretical progress in physics.

Another poster-boy of theoretical physics is certainly Albert Einstein (1879–1955). We have already mentioned his famous thought experiment of chasing a beam of light, to which Einstein later assigned a memorable role in his development of special relativity (Einstein, 1949, p. 52f). According to Einstein's recollections, this episode took place when he was only 16 years old; thus, it seems to be much easier for students to connect to this story. However, as pointed out by Norton (2013), this thought experiment does not deliver the promised result, namely, the contradiction to Maxwell's theory of electrodynamics. Furthermore, it is excluded that Einstein at the age of 16 (i.e., around 1895/96) knew Maxwell's theory already.

In Norton (2016), this episode is put into the larger context of Einstein's ostensible discovery strategies. As noted by Norton, many textbooks take issue with Einstein's remarkable creativity and promulgate what Norton calls "discovery myths" (p. 250). He singles out some of the more entrenched notions; among them are the alleged power of childish thinking and the power of breaking rules. Norton (2016) argues that by ascribing his discoveries to these "strategies" textbook authors oversimplify and distort Einstein's work grossly. For example, pointing to his apparent childlike attitude (as exemplified in the chasing-the-light-story) ignores the lengthy and painstaking preparatory work which was needed to arrive at the results. Hence, the title of Norton's paper from 2016 is "How Einstein did not discover."

10.5.2.2 The Matthew–Matilda effect

In Sec. 10.5.1.5, we have mentioned the common neglect of technicians and laboratory helpers as a consequence of the "great man history" tradition. We come back to this issue here because it is certainly not restricted to experimental practices. The above examples of Newton and Einstein illustrate again that the role of individual geniuses in science is typically emphasized—a tendency which has been (and still is) further sustained through the scientific reward system.

The focus on "great man history" is a typical trait of myth-historical narratives and Perillán (2021, p. 12) notes that while these stories "deliver value, coherence, and inspiration to their communities" they also have unintended consequences. They can "erect barriers to inclusivity, which leave women, people of color, and those with divergent ideas feeling that they donot belong" (*ibid*., p. 12).[15] The pioneering work on the systematic under-recognition of female scientists was made by Margaret Rossiter. One of the best known examples of this is the Austrian-Swedish physicist Lise Meitner, who was deprived of a share of the Nobel Prize, which was awarded exclusively to her long-time collaborator Otto Hahn (Rossiter, 1993, p. 329). The problem is, of course, that these cases are typically not well-known.[16]

About 20 years ago, Rossiter introduced the term "Matilda effect" for the systemic female under-recognition—honoring the American suffragist Matilda J. Gage, who has fallen into oblivion herself (Rossiter, 1993). The term has been coined with reference to the "Matthew effect," i.e., the over-recognition of those at the top of the scientific profession (Merton, 1968). Merton noted that when papers are coauthored by scientists of unequal reputation, the better-known author usually gets more

---

[15] Especially, the neglect of Islamic influences on the history of physics is problematic. A noteworthy exception is provided by Heilbron (2015, Chap. 2). In contrast to many histories of physics, Heilbron characterizes the Islamic science as not just preserving ancient Greek scientific knowledge but also elaborating and enriching. Many themes of later European science have been anticipated by this tradition. For example, the development of astronomy (especially in Bagdad) and the large number of observations made by Arabs would fuel European astronomical research until the 18th century.

[16] Curiously, Priemer and Lederman (2021, p. 122) confuse Lise Meitner for Marie Curie, a rather awkward attempt to display respect for a disadvantaged person.

credit. Also in cases of multiple concurrent discoveries, the credit is often misallocated to the already more famous researcher (Merton, 1968, p. 60). As a salient example, George Zweig and Murray Gell-Mann independently introduced the hypothesis of a specific nucleon substructure (called “aces” by Zweig and “quarks” by Gell-Mann; both papers were submitted in January 1964—Gell-Mann’s a bit earlier). Famously, Gell-Mann’s name prevailed and he later received the Nobel Prize (for this and other accomplishments). However, the same hypothesis was proposed even a third time. The Swiss theorist André Petermann introduced “elementary spinor particles” with all the attributes of quarks/aces in a paper submitted in December 1963 (De Rújula, 2019). His share in the quark hypothesis is almost completely forgotten.

The Matthew-effect does not only disadvantage women and strictly speaking this effect includes the under-recognition-part also, but as noted by Rossiter this meaning is less often recognized. Given that the undervaluing of women’s contributions to science is so systematic it deserves its own name (Rossiter, 1993, p. 334).[17] As recently shown by Pillion and Bergin (2022) with respect to Irish physics textbooks, this problem remains until today.

10.5.2.3 Models and modeling in physics and physics education

In physics teaching at the high school or college-level, theoretical practices are often addressed rather implicitly—so to say as the “non-experimental” activities such as problem-solving, generating hypothesis, or providing explanations based on known theories. The more advanced elements which touch upon the *generation* of theories are typically subsumed under the category of “modeling.” Hence, there is an impressive literature on models, modeling and how students are supposed to acquire “modeling skills.” This holds in general, but models are also a crucial issue in the NOS and NOSI debate [see, e.g., Lederman *et al.* (2002); Lehrer and Schauble (2006); Lederman (2007); Schwarz *et al.* (2009); and Sins *et al.* (2009)].

At the same time, the term “model” is extremely elusive—by the way, a property it shares with the term “theory.” Hence, the corresponding debate is multilayered and at times even outright confusing. In order to realize that the term “model” has no unique meaning, one only needs to compare its usage in the following examples: The standard model of particle physics, Bohr’s atomic model, the particle model for explaining the solid, liquid, or gaseous state, the model of the ideal gas, or the fluid-flow model of electricity.[18]

---

[17] It is no small irony that the term Matthew-effect was coined by the American sociologist Robert K. Merton (1968) while relying heavily on the doctoral dissertation of Harriet Zuckerman. Merton later acknowledged that she should have become co-author of the paper instead of mentioning her in the footnotes only (Rossiter, 1993, p. 334). Cases of sharing the credit unequally often concern the collaboration of spouses. Merton and Zuckerman were married in 1993.

[18] As a brief remark on these examples: The standard model is rather the currently accepted theory of particle physics, while Bohr’s model or the ideal gas qualify as theoretical models in the sense alluded below. The other examples could also be called analogies, but some (e.g., Mary Hesse) view the relation between a model and its target system as analogical (Hesse, 1972). In that case, this distinction is clearly not very helpful either.

A careful review of the model-debate in the context of science education is provided by Oh and Oh (2011). As a rough working definition, we may characterize a model as a *representation* of a "target system," e.g., objects, phenomena, processes, or ideas (Oh and Oh, 2011, p. 1113) and in general, scientific models are used in order to generate claims and to learn something about this target system (Frigg and Nguyen, 2017, p. 51).[19]

The whole model-debate is very much driven by the discussion about the meaning of the term "theory." Until the late 1960 the "received view" of theories was the so-called "Statement" or "Syntactic View," which considered theories to be axiomatized sets of sentences in a formalized symbolic language of first-order logic (Grandy, 2003). Objections against this understanding of theories centered on its axiomatic and linguistic form and it was superseded by a view that goes by many names: "Semantic View," "Non-Statement View," or "Model-Based View."[20] An influential treatment of this Model Based View (MBV) for the physics education community was due to Ronald N. Giere (1988).

According to the MBV, a theory should be viewed as a family of models and that a model is an abstract entity having all and only the properties ascribed to it by an accompanying representation. The background is the following: Giere (1988) takes classical mechanics as his paradigmatic example and his line of reasoning is developed by investigating advanced textbook accounts, i.e., he wants to characterize *science* by the way *scientific knowledge* is successfully imparted. Giere (1988, p. 76ff) first notes that, e.g., Newton's second "law" ($F = ma$) cannot be viewed as an empirical claim since it needs a specific force function to be filled in. On choosing, e.g., the linear restoring force ($F = -kx$) one describes a harmonic oscillation. However, also here the solution cannot be understood as a true universal statement about, say, all spring pendulums, since, e.g., a number of approximations had been introduced. Giere now suggests that the "harmonic oscillator" should rather be viewed as an "abstract" (and constructed) entity—a (theoretical) model for Newton's second law. Of course, there are many other types of oscillators (e.g., damped or forced). In this sense, the "oscillator" is rather a "family of models." Now, how is the connection to the "world" construed? Giere suggests that theoretical models are the means by which scientists represent the world (springs, planets or violin strings), i.e., there is a similarity relation between a model and some real system (this reflects the realistic attitude of Giere). If we put these elements together we arrive at his view on theories: A theory is identified with a family of models together with claims about the sorts of things to which the models apply (i.e., are similar to).

One reason why Giere's approach was so eagerly embraced by the physics education community is its attention paid to cognitive factors, mental models and mental representations in *doing* physics. These aspects connect nicely to central aspects of *learning* physics as well [see, e.g., Izquierdo-Aymerich and Adúriz-Bravo (2003), Crawford and Cullin (2004), or Adúriz-Bravo and Izquierdo-Aymerich (2003)].

---

[19] However, "representations" and "models" should clearly not be equated since there are many non-model-based forms of representation (e.g., lexicographical representations such as words).

[20] Still, other labels exist and Grandy (2003) suggests that these different names reflect some important uncertainty about details.

For example, Windschitl *et al.* (2008) have developed a concept of Model Based Inquiry around the generating, testing, and revising of scientific models.

However, the philosophical model-debate did not stop with the Semantic View (or its specific version by Giere). Both, the Semantic and the older Syntactic View of theories made models and theories depend on each other closely. Some of the more recent developments of the philosophical debate center around the idea that models are more theory independent and some authors argue that they may also function as semi-autonomous "mediators" between theory and experiment (Morrison, 1999; and Morrison and Morgan, 1999).[21] There are plausible and relevant examples for such models, e.g., climate models apply elements from many different theories (say, fluid dynamics, thermodynamics, and electromagnetism) cooperatively. Frigg and Hartmann (2020) conclude their discussion of this example as follows: "What delivers the results is not the stringent application of one theory, but the voices of different theories when put to use in chorus with each other in one model."

In Koponen (2007), the reception of Giere's work in physics education is summarized and it is argued that the inclusion of "models as mediators" is needed in order to provide a more robust philosophical underpinning for the model usage in physics education. This piece is part of a collection of essays in *Science & Education* on models introduced by Matthews (2007); a good starting point for exploring the various strands of the more recent model-debate in science education.

10.6 SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

There is no single Scientific Method which is common across science and only science and as David Gooding concluded even more radically: "There is no single […] philosophy of science because scientific practices are irreducibly varied" (Gooding, 1997, Sec. 122). Science is a complex epistemic and social practice which is embedded in historical and institutional contexts and to provide a fair presentation of its nature in the physics classroom is challenging. The controversy on the proper conceptualization of the Nature of Science (NOS) and the Nature of Scientific Inquiry (NOSI) within physics education is a reflection of this fact.

Consequently, our chapter could only provide a very selective overview with common misconceptions as the major organizing theme. However, as noted by Perillán, historians should be more reflective in their work and should avoid the dichotomization between "true" and "myth" history. They also tell some kind of myth-history involving necessary abridgment, selection bias and goal orientation. A similar demand can be extended to physics educators. Especially, the criteria for choosing historical case studies need more attention and physics education should draw on the similar debate in history and philosophy of science (Bolinska and Martin, 2021).

---

[21] Oh and Oh (2011, p. 1114) make the surprising claim that such a mediating function is already contained in the MBV. However, in this framework model can clearly not act as autonomous agents.

Much of the more recent work in physics education is devoted to inquiry-based teaching. As a result, some neglect of theoretical practices and methods can be observed. Exploiting case studies on the acceptance of theories along the line of Brush and Segal (2015) seems highly beneficial. In addition, there is evidence that students' science achievement is better supported by guided instruction rather than open-ended inquiry (Zhang *et al.*, 2021). These authors see the danger that science achievement is traded off by prioritizing other learning outcomes (like interest, attitude but also NOS). However, "It is hard to conceive of valid interests in and attitudes toward science without having the necessary conceptual knowledge and understanding" (Zhang *et al.*, 2021, p. 16). This is a valid concern.

REFERENCES

Adúriz-Bravo, A. and Izquierdo-Aymerich, M., Sci. Educ. **12**, 27–43 (2003).
Allchin, D., Sci. Educ. **95**, 518–542 (2011).
Allchin, D., Sci. Educ. **87**(3), 329–351 (2003).
Allchin, D., Sci. Educ. **96**, 693–700 (2012).
Berg, H. C., J. Soc. Sci. Educ. **3**(1), 25–37 (2004).
Bolinska, A. and Martin, J. D., Stud. Hist. Philos. Sci. Part A **89**, 63–73 (2021).
Boyd, N. M. and Bogen, J., *The Stanford Encyclopedia of Philosophy*, Winter 2021st ed, edited by E. N. Zalta (Stanford University, Stanford, 2021). See https://plato.stanford.edu/archives/win2021/entries/science-theory-observation/.
Brush, S. G. and King, A. L., *History in the Teaching of Physics* (University Press of New England, Hannover, NH, 1972).
Brush, S. G. and Segal, A., *Making 20th Century Science: how Theories Became Knowledge* (Oxford University Press, 2015).
Brush, S. G., Science **183**, 1164–1172, (1974).
Brush, S., *Band 8 2004*, edited by L. Danneberg *et al.* (De Gruyter, Berlin, 2021), pp. 255–264.
Chambers, D. W., Sci. Edu. **67**(2), 255–265 (1983).
Chang, H., Sci. Educ. **20**, 317–341 (2011).
Chinn, C. A. and Malhotra, B. A., Sci. Educ. **86**(2), 175–218 (2002).
CJEU (2017). Press release No 113/17, Luxembourg, Judgment in Case C-90/16, 26 October 2017.
Conant, J. B. and Nash, L. K., *Harvard Case Histories in Experimental Science* (Harvard University Press, Cambridge, MA, 1957).
Crawford, B. A. and Cullin, M. J., Int. J. Sci. Educ. **26**, 1379–1401 (2004).
Dahlin, B., Sci. Educ. **10**(5), 453–475 (2001).
Davis, E. W., Mid-West Q. **2**(1), 48–56 (1914).
De Rújula, A., Int. J. Mod. Phys. A **34**(32), 1930015 (2019).
Duhem, P., *Aim and Structure of Physical Theory* (Atheneum, New York, 1962).
Duschl, R. A. and Grandy, R., Sci. Educ. **22**(9), 2109–2139 (2013).
Duschl, R. A., *Scientific Inquiry and Nature of Science*, edited by L. B. Flick and N. G. Lederman (Springer, Dordrecht, 2006), pp. 319–330.
Duschl, R. A., *Restructuring Science Education: The Importance of Theories and Their Development* (Teachers College Press, New York, 1990).
Einstein, A., *Albert Einstein-Philosopher Scientist*, 2nd ed., edited by P. A. Schilpp (Tudor Publishing, New York, 1949), Vol. 1951, pp. 2–95. pp. 671–72. Reprinted with a correction as Autobiographical Notes. La Salle and Chicago Open court, 1979.
Elby, A. and Hammer, D., Sci. Educ. **85**, 554–567 (2001).
Emden, M. and Gerwig, M., Sci. Educ. **29**, 589–616 (2020).
Erduran, S. and Dagher, Z. R., *Contemporary Trends and Issues in Science Education* (Springer, Dordrecht, 2014), Vol 43.
Fara, P., *Newton's Apple and Other Myths About Science*, edited by R. L. Numbers and K. Kampourakis (Harvard University Press, Cambridge, MA, 2015), pp. 48–56.
Feynman, R., *QED: The Strange Theory of Light and Matter*, 3rd ed. (Princeton University Press, Princeton, NJ, 2006).
Franklin, A., Phys. Perspect. **18**, 3–57 (2016).
Franklin, A., *The Neglect of Experiment* (Cambridge University Press, 1989).
Frigg, R. and Hartmann, S., *The Stanford Encyclopedia of Philosophy*, Spring 2020 Edition, edited by E. N. Zalta (Stanford University, Stanford, 2020).
Frigg, R. and Nguyen, J., *Springer Handbook of Model-Based Science*, edited by L. Magnani, and T. Bertolotti (Springer, Dordrecht, 2017), pp. 49–102.
Gehrenbeck, R. K., Phys. Today **31**, 34–41 (1978).
Giere, R., *Explaining Science: A Cognitive Approach* (University of Chicago Press, Chicago, IL, 1988).
Gieryn, T. F., Am. Sociol. Rev. **48**(6), 781–795 (1983).
Gieryn, T. F., *Cultural Boundaries of Science: Credibility on the Line* (University of Chicago Press, Chicago, IL, 1999).

Gooding, D., Isis **88**(1), 121–122 (1997).
Gordin, M. D., *Newton's Apple and Other Myths About Science*, edited by R. L. Numbers and K. Kampourakis (Harvard University Press, Cambridge, MA, 2015), pp. 219–226.
Grandy, R. and Duschl, R., Sci. Educ. **16**, 141–166 (2007).
Grandy, R. E., Sci. Educ. **42**, 773–777 (2003).
Grebe-Ellis, J. and Passon, O., Dialogue **1**, 50–59 (2020).
Gyllenpalm, J. and Wickman, P.-O., Sci. Educ. **95**, 908–926 (2011).
Hall, A. R., Hist. Sci. **21**(1), 45–59 (1983).
Heering, P. and Höttecke, D., *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. Matthews (Springer, Dordrecht, 2014).
Heering, P., Sci. Educ. **9**(4), 363–373 (2000).
Heering, P., Am. J. Phys. **60**, 988–994 (1992).
Heering, P., *Restaging Coulomb: Usages, Controverses et Réplications Autour de la Balance de Torsion*, edited by C. Blondel and M. Dörries (Leo S. Olschki, Firenze, 1994), pp. 47–66.
Heering, P., *Lavoisier in Perspective*, edited by M. Beretta (Deutsches Museum, München, 2005), pp. 27–41.
Heilbron, J. L., *Physics: A Short History From Quintessence to Quarks* (OUP, Oxford, 2015).
Hentschel, K., Phys. Z. **34**, 225–231 (2003).
Hepburn, B. and Andersen, H., *Scientific Method. The Stanford Encyclopedia of Philosophy*, Summer 2021st Edition, edited by E. N. Zalta (Stanford University, Stanford, 2021).
Hesse, M., *The Encyclopedia of Philosophy*, edited by P. Edwards (MacMillan, New York, 1972), pp. 354–359, Vol. 5.
Höttecke, D. and Silva, C. C., Sci. Educ. **20**(3-4), 293–316 (2011).
Höttecke, D., 12th IHPST Conference, Pittsburgh, PA (2013). Retrieved from http://archive. ihpst. net/2013/Procs/Hottecke% 20symposium.pdf.
Höttecke, D. *et al.*, Sci. Educ. **21**, 1233–1261 (2012).
Irzik, G. and Nola, R., Sci. Educ. **20**(7–8), 591–607 (2011).
Izquierdo-Aymerich, M. and Adúriz-Bravo, A., Sci. Educ. **12**, 27–43 (2003).
Jiang, F. and McComas, W. F., Int. J. Sci. Educ. **37**(3), 554–576 (2015).
Kang, S. *et al.*, Sci. Educ. **89**, 314–334 (2004).
Klein, M. J., *The Use and Abuse of Historical Teaching in Physics* (Brush and King, 1972), pp. 12–18.
Klopfer, L. E., *History of Science Cases (HOSC)* (Science Research Associates, Chicago, IL, 1964).
Koponen, I. T., Sci. Educ. **16**, 751–773 (2007).
Kuhn, T. S., *The Structure of Scientific Revolution*, Third edition (The University of Chicago Press, Chicago, IL, 1996).
Latour, B., *The Social Dimensions of Science*, edited by E. McMullin (Notre Dame University Press, Notre Dame, 1992), pp. 272–292.
Laudan, L., *Physics, Philosophy and Psychoanalysis: Essays in Honor of Adolf Grünbaum*, edited by R. S. Cohen and L. Laudan (D. Reidel, 1983), pp. 111–127.
Laudan, L., *Science and Hypothesis* (D. Reidel, Dordrecht, 1981).
Laugksch, R. C., Sci. Educ. **84**(1), 71–94 (2000).
Lederman, J. S. *et al.*, J. Res. Sci. Teach. **51**, 65–83 (2014).
Lederman, L. and Teresi, D., *The God Particle: If the Universe is the Answer, What is the Question?*, 2nd ed. (Mariner Books, New York, 2006).
Lederman, N. G. and Lederman, J. S., Discip. Interdiscip. Sci. Educ. Res. **1**(1), 1–9 (2019).
Lederman, N. G., *Handbook of Research on Science Education*, edited by S. K. Abell and N. G. Lederman (Mahwah: Lawrence Erlbaum Associates, 2007), pp. 831–880.
Lederman, N. G. *et al.*, J. Res. Sci. Teach. **39**(6), 497–521 (2002).
Lehrer, R. and Schauble, L., *Handbook of Child Psychology*, 6th ed., edited by W. Damon *et al.* (Wiley, 2006), Vol. 4.
Matthews, M. R., Interchange **20**(2), 3–15 (1989).
Matthews, M. R., *Advances in Nature of Science Research*, edited by M. Khine (Springer, Dordrecht, 2012).
Matthews, M. R., Sci. Educ. **1**(1), 11–47 (1992).
Matthews, M. R., Sci. Educ. **16**, 647–652 (2007).
McComas, W. F. and Clough, M. P., *Nature of Science in Science Instruction*, Science Philosophy, History and Education, edited by W. McComas (Springer, Cham, 2020).
McComas, W. F. and Olson, J. K., *The Nature of Science in Science Education: Rationales and Strategies*, edited by W. F. McComas (Kluwer, Dordrecht, 1998), pp. 41–52.
McComas, W. F., *Adapting Historical Knowledge Production to the Classroom*, edited by P. V. Kokkotas *et al.* (Sense Publishers, Rotterdam, 2011), pp. 37–53.
Merton, R. K., Science **159**(3810), 56–63, (1968).
Metz, D. *et al.*, Sci. Educ. **16**, 313–334 (2007).
Meyer, K. and Carlisle, R., Int. J. Sci. Educ. **18**(2), 231–248 (1996).
Middleton, W. E. K., *The History of the Thermometer and its Use in Meteorology* (Johns Hopkins University Press, Baltimore, MA, 1966).
Mizrahi, M., Int. Stud. Philos. Sci. **29**(2), 129–148 (2015).
Morrison, M. and Morgan, M., *Models as Mediators*, edited by M. S. Morgan and M. Morrison (Cambridge University Press, Cambridge, MA, 1999).

Morrison, M., *Models as Mediators*, edited by M. S. Morgan and M. Morrison (Cambridge University Press, Cambridge, 1999), p. 21.
Norton, J. D., *Thought Experiments in Philosophy, Science and the Arts*, edited by J. R. Brown *et al.* (Routledge, New York, 2013), pp. 123–140.
Norton, J. D., Phys. Perspect. **18**, 249–282 (2016).
Oh, P. S. and Oh, S. J., Int. J. Sci. Educ. **33**(8), 1109–1130 (2011).
Osborne, J. F. *et al.*, J. Res. Sci. Teach. **40**(7), 692–720 (2003).
Østergaard, E. *et al.*, Stud. Sci. Educ. **44**(2), 93–121 (2008).
Park, W. and Song, J., Sci. Educ. **27**, 39–61 (2018).
Passon, O. and Grebe-Ellis, J., Eur. J. Phys. **38**(3), 035404 (2017).
Perillán, J. G., *Science Between Myth and History: The Quest for Common Ground and Its Importance for Scientific Practice* (Oxford University Press, 2021).
Pigliucci, M., *Philosophy of Pseudoscience: Reconsidering the Demarcation Problem*, edited by M. Pigliucci and M. Boudry (University of Chicago Press, Chicago, 2013), pp. 9–28.
Pillion, K. and Bergin, S. D., Phys. Educ. **57**, 025017 (2022).
Popper, K., *The Logic of Scientific Discovery* (Routledge, Abingdon-on-Thames, 2002) [1959].
Priemer, B. and Lederman, N. G., *Physics Education* (Springer, Cham, 2021), pp. 113–150.
Quine, W. V. O., *From a Logical Point of View* (Harvard University Press, Cambridge, MA, 1953).
Rang, M. and Grebe-Ellis, J., J. Gen. Philos. Sci. **49**, 515–523 (2018).
Ribe, N. and Steinle, F., Phys. Today **55**(7), 43–49 (2002).
Rossiter, M. W., Soc. Stud. Sci. **23**, 325–341 (1993).
Salchegger, S. *et al.*, Large-scale Assess. Educ. **9**(14), 23 (2021).
Schwartz, R. S. *et al.*, Paper Presented at the International Conference of the National Association for Research in Science Teaching (NARST), Baltimore, MD (2008).
Schwarz, C. V. *et al.*, J. Res. Sci. Teach. **46**(6), 632–654 (2009).
Seroglou, F. and Koumaras, P., *Science Education and Culture*, edited by F. Bevilacqua *et al.* (Springer, Dordrecht, 2001).
Shapin, S., Am. Sci. **77**(6), 554–563 (1989).
Sins, P. H. M. *et al.*, Int. J. Sci. Educ. **31**(9), 1205–1229 (2009).
Sommer, W., *Phase Transition Water/Steam—Physics 9th Grade* (Pädagogische Forschungsstelle Kassel, Kassel, 2020), Vol. **2**.
Sommer, W., *Physics 10th Grade—Statics—Kinematics—Dynamics* (Pädagogische Forschungsstelle Kassel, Kassel, 2022).
Sommer, W., *Physics in Waldorf Schools* (Pädagogische Forschungsstelle Kassel, Kassel, 2021).
Steiner, R., *Goethean Science* (Mercury Press, New York, 1988).
Steinle, F., *Revisiting Discovery and Justification*, edited by J. Schickore and F. Steinle (Springer, Dordrecht, 2006), pp. 183–195.
Steinle, F., Philos. Sci. **64**, S65–S74 (1997).
Stinner, A. *et al.*, Sci. Educ. **12**, 617–643 (2003).
Teig, N., Scherer, R., and Nilsen, T., Learn. Instrum. **56**, 20–29 (2018).
Thurs, D. P., *Newton's Apple and Other Myths About Science*, edited by R. L. Numbers and K. Kampourakis (Harvard University Press, Cambridge, MA, 2015), pp. 210–218.
Tobin, K. and McRobbie, C. J., Sci. Educ. **6**, 355–371 (1997).
Whitaker, M. A. B., Phys. Educ. **14**, 108–112 (1979). (part 1) and pp. 239–242 (part 2).
Windschitl, M. *et al.*, Sci. Educ. **92**(5), 941–967 (2008).
Wong, S. L. and Hodson, D., Sci. Educ. **93**(1), 109–130 (2009).
Zhang, L. *et al.*, Educ. Psychol. Rev. **34**(2), 1157–1176 (2021).

CHAPTER 11

EPISTEMIC BELIEFS AND PHYSICS TEACHER EDUCATION[1]

Gábor Á. Zemplén

Zemplén, G. Á., "Epistemic beliefs and physics teacher education," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 11-1–11-30.

11.1 INTRODUCTION

Children's cognitive maturation testifies to increasingly complex webs of beliefs and skills, and learners gradually acquire epistemologically significant beliefs. Some years after *having beliefs*, *t*he object of research emerges, and one can talk of epistemic beliefs, beliefs related to the nature of learning and knowledge-creation. Physics teachers' target audience has certain epistemic beliefs, and a major driving force of research is whether these beliefs relate to academic achievement, and how instruction can be tailored to provide conducive learning environments if we understand the student population's epistemic beliefs.

The topic links the study of children's cognitive development with life-long learning, and the potential to achieve a coherent epistemology. Research focus on epistemic beliefs was historically most closely related to studying undergraduates, and developmental research studied maturation, finding that a well-structured system of epistemic beliefs before early adulthood is atypical. Accordingly, teacher-trainees as well as Ph.D. students can be considered to be on their way to acquire mature and consistent epistemological views.

This chapter introduces select and venerable traditions of research. Section 11.2. discusses some of the historical trajectories, but the broader perspective could show how adaptation of EB research was influenced by changes in didactics globally, while local research was influenced by particular

---

[1] The helpful comments by the reviewers as well as by Douglas Allchin, Csaba Pléh and Benedek Láng are acknowledged.

https://doi.org/10.1063/9780735425514_011

psychological assumptions about development (e.g., Kohlberg, Piaget, and Vygotsky, 1978), showing many variations in the various countries. Research on science education itself evolved shifting focus from knowledge learning, identity, and informal education to an increased emphasis on scientific literacy, socio-science issues (SSI) and formal education, and most recently an emphasis on scientific argumentation and STEM education (for the grouping of the last two decades of research, see Wang *et al.*, 2022). The section focuses on classics of the developmental tradition to show how an ever-finer grained approach took hold, increasing the dimensionality of research as well as the domain-specific uses. Dimensional research utilized some tenets of developmental studies and introduced new assumptions. Narrowly understood epistemic (or epistemological) beliefs (EB) are multidimensional artificial constructs that can be used to ascertain and measure change in epistemic beliefs, where change is usually investigated in relation to instruction, student achievement, and scores on various other measurement tools.

It should be noted at the outset that repeated attempts have been made to "carve out" a well-defined conceptual space for epistemic beliefs. The strands of research have all benefited from disciplines ranging from developmental psychology and philosophy, but the impact of the pedagogical research on these areas remains limited. Section 11.3 addresses foundational questions and highlights some notable issues concerning the methodology, the philosophical basis, and contradictory assumptions behind some of the commonly used measuring techniques. While there is a lack of consensus on the definition and a lack of consensus on use in the field, this area has yielded an ever-increasing empirical knowledge-base. Without decades of continued research in this field, we would not know the significance of the challenges and trade-offs that the research on "epistemic beliefs" poses. To introduce the field for the reader, I provide short descriptions by highlighting examples that help map the strands of research as well as the epistemological and methodological issues, and introduce some key issues that are unsettled despite decades of research. These appear necessary as the field is "ill-structured" in comparison to physics, and with a general desire to produce novelty in the study of the interactions between EBs and performance, the proposed frameworks have been diverging. Section 11.4. discusses some of the future avenues of research on epistemic beliefs.

11.2 FROM THE EMERGENCE OF RESEARCH ON EPISTEMIC BELIEFS TO PROLIFERATION

To outline a historical trajectory of methodological innovation, this section provides a highly selective introduction to several of the approaches, on which later sections (discussing foundational issues and future trends) also build. The new methodologies introduced can be considered as major turning points, as all increased the number of meaningful research questions and resulted in a plethora of employed instruments, many maintained, modified, and fit to specific purposes or localities up to today.

Studying cognitive development involves studying the increasing complexity of a cognizing agent; therefore, I introduce Fischer's hierarchical approach (Sec. 11.2.1) before moving on to developmental accounts of the sophistication of epistemic beliefs. An early and influential source is the dualist-relativist continuum described by Perry (1970), another is the "Reflective Judgment" model by Kitchener and King (Secs. 11.2.2 and 11.2.3). These long-term developmental studies have been used to ground multidimensional approach by Schommer, and today an increasing ratio of the studies define and adopt the construct of EB as multidimensional, where the dimensions are usually assumed to be on a naïve-sophisticated continuum (Sec. 11.2.4).

While earlier approaches tended to focus on development, responding to growth in research in didactics, applications within domains and subdomains (physics, and further divisions within physics) increasingly take hold. The last subsections introduce the increased focus on *domain-specific* investigations, and the quest to find the right resolution to study learning events and map physics-specific conceptual change (Secs. 11.2.5 and 11.2.6).

11.2.1 Hierarchies of skills (Fisher)

Skill-theory (Fisher, 1980) is a well-known and straight-forward approach to assume a hierarchy (of skills) that appears in the course of the development starting at birth. Fisher's approach has the benefits of explicating some *assumptions about the levels of complexity* to be ascribed to research subjects at a specific age. This increasing complexity (see Table 11.1) has important implications for learning (e.g., manipulating numbers, equations, or tackling non-algorithmic problems), and can provide a guideline to assess necessary conditions for having beliefs, having epistemic beliefs, or having well-structured systems of epistemic beliefs. Functional performance for Fisher could increase until mid-adulthood (Fischer *et al.*, 2003).

Fisher's framework had specific assumptions about cognitive functioning, starting as reflexes, followed by actions, representations, and abstractions, and also about the "elementary operations" to be employed: "cognitive skills can be described effectively and precisely in terms of elementary set theory" (Fisher, 1980, p. 481). It could be used to study the emergence of (previously non-existent) understanding, and Piaget-inspired examples (like the conservation of length of chord) helped model the construction of a new understanding (p. 484). Fisher provided little argumentation for the number of levels: the number "seemed" adequate. However, based on the model, he predicted that *once a new level is attained, there will be "spurts of development" as opposed to a monotonic increase* (p. 485). To explain developmental transitions, Fisher's work continued to tackle conceptual issues of dynamical development (for the use of non-linear dynamics in a developmental framework, see Rose and Fisher, 1998). In learning theory, similar hierarchical taxonomies were proposed, like Structure of the Observed Learning Outcomes (SOLOs) or Bloom's taxonomy, also influencing course descriptions.

Table 11.1

Fisher's hierarchy of skills: Age Periods at Which Levels First Develop (Fisher 1980, p. 522), with the last two columns on optimal and functional developmental levels during the school years (Fisher 2008, p. 133).

Cognitive level	Age period[a]	Optimal	Functional[b]
1: Single sensory-motor sets	Several months after birth		
2: Sensory-motor mappings	Middle of the first year		
3: Sensory-motor systems	End of the first year and start of second year		
4: Systems of sensory-motor systems, which are single representational sets	Early preschool years	2 years	2 to 5 years
5: Representational mappings	Late preschool years	4	4 to 8
6: Representational systems	Grade school years	6	7 to 12
7: Systems of representational systems, which are single abstract sets	Early high school years	10	12 to 20
8: Abstract mappings	Late high school years	15	17 to 30
9: Abstract systems	Early adulthood	20	23 to 40[c]
10: Systems of abstract systems	Early adulthood	25	30 to 45

[a] These periods are merely estimates for middle-class Americans. For Levels 9 and 10, existing data do not allow accurate estimation.

[b] Ages for functional levels vary widely and are coarse estimates, based on research by Dawson, Fischer, Kitchener, King, Kohlberg, Rest, and others. Levels are highly related to education.

[c] Or never, for many domains.

11.2.2 Evaluating knowledge claims and controversial issues (Kitchener–King)

With a surge in research in the 1980–1990s, various developmental models were proposed, where the endpoint, after the gradual refinement of critical thinking, was enabled by the increasingly complex representations and systems. A *seven-stage model of post-adolescent reasoning* styles in King *et al.* (1983) gave an account of the (presumably highly complex) monitoring that is involved when older adolescents and adults are faced with ill-structured problems (see also, Kitchener and King, 1981). The problems about which rational people reasonably disagree are increasingly relevant for science education, as they include ecological and health risks.

The *Reflective Judgment Model* touched on the Deweyan tradition concerned with developing democratic/citizenship skills and teaching about the nature of science, including ill-structured problems. It "describes a developmental progression that occurs between childhood and adulthood in the ways that people understand the process of knowing and the certainty of knowledge claims and in the corresponding ways that they justify their beliefs." (King, 2000, p. 37). As with many in the emerging wave of the cognitive revolution (leaving behind neo-behaviorist theories), the authors were critical towards but inspired by Piaget's genetic epistemology. Piaget's theory of cognitive development focused on earlier stages of development, and when tackling a theory of adult epistemologies, Kitchener raised the issue of empirical testability (for an interesting analysis of testing Galileo's laws and Piaget's theory, see Kitchener, 1993).

The framework emerged after a 10-year longitudinal study of three different cohorts, and its basis was the so-called *Reflective Judgment Interview* (over 1700 participants), interviews with individuals about ill-structured problems. After probing the reasoning of participants under non-optimal circumstances, responses were analyzed and scored. Three levels of increasing complexity were distinguished (King *et al.*, 1983):

- Level 1—cognitive processing: individuals engage in processes like computing, memorizing, reading, and perceiving.
- Level 2—meta-cognitive processing: ability to monitor one's progress when engaged in level one tasks.
- Level 3—epistemic cognition: "the limits of knowing, the certainty of knowing, and the criteria for knowing."

King and Kitchener have argued that epistemic cognition is utilized when individuals are engaged in ill-structured problem solving (King and Kitchener, 1994). While Level 1 and Level 2 processes are monitored by Level 2 processing, *the foundation of critical thinking is epistemic cognition* that can *monitor the epistemic nature of problems* and problem types (are they solvable?) and to evaluate proposed (non-algorithmically solvable) solutions.

The three major periods were further split into stages: the prereflective (Stages 1–3), the quasi-reflective (Stages 4 and 5), and the reflective (Stages 6 and 7). Reference to seven distinct but developmentally related sets of assumptions about the process of knowing (view of knowledge) and how it is acquired (justification of beliefs) assumed *a scale of development, where each successive set of epistemological assumptions is characterized by a more complex and effective form of justification* (King and Kitchener, 1994, see also Table 11.2). The results demonstrated that post-adolescents' conceptions of justification change over age/educational levels. Highly significant differences were found between age/educational groups (Kitchener and King, 1981).

The authors stressed that although neatly distinct levels are suggested to allow scoring, *individuals are not "in" a single stage or "at" a certain level.* As variability of stage reasoning was common (that is, evidence of reasoning that is characteristic of more than one stage at a time), the mixture of scores led King, Kitchener, and Wood (1994) to assume that there is a plurality within the individual and to characterize development as the following:

> "…waves across a mixture of stages, where the peak of a wave is the most commonly used set of assumptions. While there is still an observable pattern to the movement between stages, this developmental movement is better described as the changing shape of the wave rather than as a pattern of uniform steps interspersed with plateaus." (King *et al.*, 1994, p. 140)

Table 11.2

Major periods and sample responses of the *Reflective Judgment Model* after King and Kitchener (1994).

Major period and characteristics	Typical response for specific stages
Prereflective Reasoning Trusting the word of an authority figure or through firsthand observation rather than, for example, through the evaluation of evidence.	1 "I know what I have seen." 2 "If it is on the news, it has to be true." 3 "When there is evidence that people can give to convince everybody one way or another, then it will be knowledge, until then it is just a guess.'
Quasi-Reflective Reasoning They recognize that knowledge—or, more accurately, knowledge claims—contain elements of uncertainty, which they attribute to missing information or to methods for obtaining the evidence. Although they use evidence, they do not understand how evidence entails a conclusion	4 "I'd be more inclined to believe evolution if they had proof It's just like the pyramids: I don't think we'll ever know. Who are you going to ask? No one was there." 5 "People think differently and so they attack the problem differently. Other theories could be as true as my own, but based on different evidence."
Reflective Reasoning They believe they must actively construct their decisions, and that they must be evaluated in relation to the context in which they were generated to determine their validity.	6 "It is very difficult in this life to be sure. There are degrees of sureness. You come to a point at which you are sure enough for a personal stance on the issue." 7 "One can judge an argument by how well thought-out the positions are, what kinds of reasoning and evidence are used to support it, and how consistent the way one argues on this topic is as compared with other topics."

People's epistemic assumptions change over time in a developmental fashion from early adolescence to adulthood, and the supposition is that "at their best" fully grown humans are potentially reflective and reasonable agents. The expectations of the target state had their origin in Dewey (1933), but what the research suggested was rather pessimistic: "the developmental levels of college graduates probably will not be sufficient for the kinds of problems they will be asked to address in a myriad of adult roles." (*ibid.*)

The authors were clearly interested in the foundational issues that surround the analyzability of the epistemic subject, but the kind of thinking referred to as reflective judgment is not easily operationalized. The model scenarios to study reflective judgment involved critical thinking about ill-structured problems and in the interview situations, the judgments on the adequacy of knowledge claims made had to be tentative (reasonably certain), and in need of re-evaluation if new data or new methodologies become available. Reflective Reasoning is the stage where knowledge claims are not made with certainty but are also not immobilized by the admission of uncertainty (for an example of treating uncertainty as an asset for science education, see Kampourakis and McCain, 2019). Although the aim of the *Reflective Judgment* interviews was to *develop an objectively scored measure*, the specific scenario for the investigation tended to underestimate the cognitive abilities of the students. Designed for eliciting skills used in "everyday" functioning, it did not guarantee peak performance (Kitchener and Fischer, 1990; and Kitchener *et al.*, 1993). To reach beyond the typical modes of functioning and respond at higher developmental levels is possible under optimal conditions in a conducive environment with contextual feedback and emotional support. Both Fisher and Kitchener and King (in later sections referred to as K&K) proposed theories where in later stages of development there might be *large differences between functional performance* elicited using typical self-report methods or interviews *and optimal performance* in high support contexts.

11.2.3 Growing up to face relativism (Perry)

Perry began studying personality differences in (mostly male) Harvard freshmen in another influential longitudinal and qualitative study. Successful learning did not seem to depend only on motivation and study skills but also on how knowledge was viewed. His work yielded a roadmap, often portrayed as the dualist-relativist continuum: while first-year college students often have a dualistic view of knowledge (and expect the "truth" from lecturers, or from reading books) with the years they have an evolving capacity for intellectual commitments and a growing tendency to endorse positions in the face of relativism, with a devolving tendency to blind obedience to (epistemic) authority (Perry, 1970).

Perry, among others, analyzed selected participants' responses at the end of each academic year to the question "Would you like to say what has stood out for you during the year?," and discerned *nine distinct stages or "positions"* from which to view the world. In Perry's original conceptualization of the scheme, the first and the last positions were hypothetical extensions extrapolated from the empirical work. Positions 1 through 5 describe the cognitive development, intellectual refinement toward increasing differentiation and complexity. Positions 6 through 9 describe the ethical maturation (Table 11.3).

Theoretically, it is interesting that for both Perry and K&K, the process of sophistication implies a more refined evaluation of authorities, of relativism and types of uncertainties, finally transgressing to context-dependent decision-making. Complexity increases throughout the development in K&K's RJI and also in Fisher's stages, where the last stage "Systems of abstract systems" is renamed in later writings as "Single

Table 11.3
Perry's positions and encountered challenges after Perry (1970).

Major categories/positions in sequence	Encounters with diversity/multiples
Dualism (1, 2)	Multiple opinions about a given subject or issue (Positions 1 through 3)
Multiplicity (3, 4) Contextual Relativism (from 5)	Multiple contexts/perspectives from which to understand or analyze issues or arguments (Positions 4 through 6)
Commitment within Relativism (6–9)	Multiple Commitments through which one defines his or her values and identity (Positions 7 through 9)

principles." For Perry, however, once Position 5 there are no structural changes. Conversely, the later positions take ethical considerations, such as commitment, into account, but the earlier ones do not.

The Perry scheme is evaluative, the *meaning-making shifts* outline a unique developmental path after a confrontation with levels of "multiplicity," as pluralism with respect to new learning is gradually engrained. The original research targeted liberal arts education, but his scheme and the ones inspired by it were used to "measure" epistemic beliefs for students of all domains. Learners confront the university environment and the educational process shapes what students think learning is about (Ramsden, 1988; and Fry *et al.*, 2008). The *naïve—sophisticated continuum*, however, provides a conceptual framework where the epistemological belief that knowledge is increasingly conjectural and uncertain is considered a productive insight in all domains, from studying Newtonian mechanics to learning about climate change.

11.2.4 Increasing dimensionality of the object of research (Schommer)

Marlene Schommer is recognized as a major innovator of written instruments assessing personal epistemology and her *Epistemological Beliefs Questionnaire* increased the dimensionality of research. Schommer's take on personal epistemology was to assume that it is *not* unidimensional but a *system of beliefs* that were presumably *strongly linked to classroom learning and performance*. These dimensions were assumed to be independent dimensions possibly orthogonal to one another. In the frameworks introduced earlier, the development of personal epistemology was mapped as unidimensional, temporal progression through various stages, perspectives, etc., but Schommer assumed the existence of a system:

> By system, I mean that there is more than one belief to consider. And by almost independent, I mean that individuals may be sophisticated in some beliefs but not necessarily sophisticated in other beliefs. With this conceptualization, epistemological beliefs can be studied individually or in various combinations. An underlying assumption is that individual beliefs as well as unique combinations of beliefs may have different effects on learning. For example, individuals who believe in absolute (certain) knowledge that is simple (compartmentalized) may study history by memorizing lists of facts and dates. Furthermore, they may assume that all historical information is objective. On the other hand, individuals who believe in absolute knowledge that is highly complex (interconnected) may search for the big picture and relate events to each other. (Schommer, 1994, p. 300)

Schommer's work was closely linked to academic learning, beliefs in the speed, and control of knowledge acquisition (1990), but its central contribution was to respond to complexity by carving up the problem-space. Her research initiated an empirical strand of investigation that could give a finer grained analysis by separating beliefs in the source, certainty, and organization of knowledge. As most early studies, it was carried out on college students, but it was not a philosophically rooted project, like Kitchener's stages of Reflective Judgment, assuming some quality of the citizens both acquirable and

Table 11.4

Common dimensions for epistemological beliefs (stemming from Schommer's work).

Dimension	Naïve	Sophisticated
Source of Knowledge (from omniscient source to empirically evidenced-based nature of knowledge)	Knowledge comes from authority, e.g., textbooks	Knowledge is derived from reasoning/thinking/testing
Certainty of Knowledge/the stability of knowledge (from factual to constantly changing nature of knowledge)	Knowledge is not changing over time, it is fixed, absolute.	Knowledge is constantly evolving changing over time
Simplicity of knowledge/the structure of knowledge (from simple to complex nature of knowledge)	Knowledge is compartmentalized / accumulation of facts	Knowledge is highly integrated, concepts are interrelated, knowledge is relative, contingent and contextual
Innate ability/the ability to learn (from fixed or innate to incremental nature of ability).	Ability to learn is genetically pre-determined	Ability to learn is acquired by experience
Speed of Knowledge Acquisition/the speed of learning (from quick to gradual nature of learning)	Learning is quick	Learning is a gradual process

desirable. With a shift from the epistemological and developmental research questions, the approach was innovative for the learning sciences, focusing on the educational relevance of beliefs about the nature of knowledge and the orientation to the learning process.

Building on Perry's positions, King and Kitchener (and other research, e.g., Ryan's dualistic scale) developed a questionnaire using various statements extracted from the literature and grouping them into subsets, and hypothetical epistemological beliefs. The *five beliefs were each made up of two or more subsets, with each subset consisting of short statements*. So, for example, a hypothetical epistemological belief was "Knowledge is certain rather than tentative" labelled "Certain knowledge," its two subsets were "Knowledge is certain" and "Avoid ambiguity," with statements like "Scientists ultimately get to the truth," and "I don't like movies that don't have an ending." Table 11.4 gives an overview (collated from common uses of the research).

"Simple Knowledge" was somewhat overrepresented (19 question), and the statements were rated on a Likert scale by the respondents from 1 (strongly disagree) to 5 (strongly agree). Using the 12 subsets of items as variables (and not the 63 items themselves), the factor analysis reported in these studies opened up a promising line of research to find the appropriate dimensionality of the object of research, and to study the effect of epistemological beliefs in order to uncover the multiple links between personal epistemology and various aspects of learning. While earlier models used one scale to measure epistemological development, multi-dimensional research operates with several scales. As the beliefs do not necessarily develop in synchrony, a more nuanced picture can be acquired by increasing dimensionality.

With the spread of multi-dimensional research carried out via questionnaires as opposed to in-depth interviews and time-consuming longitudinal studies, the number of publications quickly grew.

Schommer was early on interested in the domain-specificity of epistemological beliefs (Schommer and Walker, 1995). Subsequent research has increasingly focused on discipline-specific beliefs, often reducing the scope (to one or a few beliefs) but investigating beliefs in more detail.

11.2.5 Domain-specificity and epistemological beliefs in physics

The models introduced so far did not focus on the sciences and paid no special attention to the study of physics. This tradition of research into personal epistemology initially assumed that the respondent has some theory of knowledge (more or less implicit, but the responses can explicate it) and that beliefs vary little depending on the subject matter, presupposing that epistemological beliefs are domain-general. Once this assumption is set aside, new research questions were raised, and several new research tools were developed. For an overview of physics-related research tools, https://www.physport.org/assessments/ is a good resource, and here are some mentioned in the chapter:

Maryland Physics Expectation survey (MPEX), Redish *et al.* (1998)
Views About Science Survey (VASS), Halloun and Hestenes (1998)
Epistemological Beliefs Assessment for Physical Sciences (EBAPS), Elby (2001)
Colorado Learning Attitudes about Science Survey (CLASS), Adams *et al.* (2006)

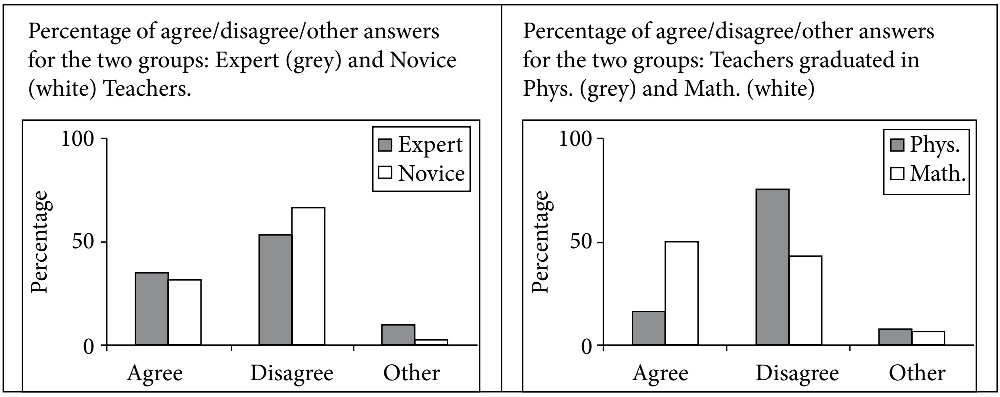
We should note that there are many adaptations to national curricular demands; for example, the Oldenburg Epistemic Beliefs Questionnaire (OLEQ) was developed from the Epistemic Beliefs Inventory (EBI). The rich cross-fertilization of fields does not make it easy to provide an overview. There are questionnaires where some items show close similarity to items in the Nature of Science (NOS) questionnaires or to consensus-statement items. For specific research designs, only parts of questionnaires are used, or some of the items are dropped for the confirmatory factor-analysis. Therefore, instead of a technical overview, I introduce some of the newer strands of research with exemplars of the types of inquiries that are relevant for physics education.

Looking for *correlations of specific epistemological beliefs with learning in specific disciplines* is a potential road to study. Using appropriate research constructs, e.g., questionnaires on the nature of science, such differences can be nicely illustrated. Sperandeo-Mineo (2012) studied responses of 127 novice teachers (2–8 years of experience) and 97 teachers with substantial experience (11–25 years) and found significant differences between the respondents of the groups on specific statements: the *position of teacher changes with teaching experience and is also influenced by the nature of their degree* (Mathematics or Physics, see Fig. 11.1.).

Another fruitful approach is to assume that *individuals are likely to hold differing beliefs about disciplines* (Hofer, 2000). One simple way is to use stock sentences already tested in a modified form; for example, instead of "sciences" one can use "physics" or "mathematics." The Maryland Physics Expectations Survey (MPEX); Redish *et al.* (1998), for example, includes the statement "Knowledge in physics consists of many pieces of information, each of which applies primarily to a specific situation," but

Statement $N°$ 3

Scientific work is a logical process, rather than intuitive and creative



Percentage of agree/disagree/other answers for the two groups: Expert (grey) and Novice (white) Teachers.

Percentage of agree/disagree/other answers for the two groups: Teachers graduated in Phys. (grey) and Math. (white)

The χ^2 test did indicate a difference between the groups at a level of $p < 0.05$

The χ^2 test did indicate a difference between the groups at a level of $p < 0.001$

FIG. 11.1

Views about knowledge-production are affected by both domain (Mathematics and Physics), and time of immersion in a field (below or above 8 years) from Sperandeo-Mineo (2012).

another domain could also be named. One can also mix statements and include domain specific beliefs, for example, in a more recent study, exploratory factor analysis yielded mathematical and general epistemological belief factors (Schommer-Aikins *et al.*, 2013).

In *domain-specific literature*, it is often found that performance correlates better with domain specific epistemological beliefs than with general epistemological beliefs, suggesting that better resolution of the research tool generally provides a more reliable map of the students' beliefs relevant for learning. These beliefs are not necessarily consistent, but usually better academic performance comes with higher consistency between domain general and domain specific epistemological beliefs. Lower academic performance just as a lower level of immersion in a domain (less content knowledge) usually implies less consistency in beliefs.

If "science" can be split into specific disciplines to obtain a more refined construct, so can "physics," acknowledging that neither "science" nor "physics" is homogenous. *Physics-specific epistemic beliefs* can further be broken down and can be studied at a sub-domain level. For example, using the Colorado Learning Attitudes about Science Survey (CLASS), selected items were turned into two items (Dreyfus *et al.*, 2019), instead of "physics" in the modified items "classical physics" and "quantum physics" were used. In modern physics and quantum mechanics courses for engineering and physics students, the "split" results indicate that

> "…most students do perceive epistemological differences to exist between classical and quantum physics, and the interview analyses suggest that at least some of these epistemological splits correspond to students' held views. Therefore, when we talk about students' epistemologies of physics, we need to be careful about treating 'physics' as homogeneous; these results show that a significant number of individual students display different approaches to knowledge in physics depending on the specific sub-discipline within physics. This exploratory study does not invalidate the CLASS, nor does it create an alternative survey instrument. Rather, it provides a proof of concept that students' epistemological views may differ by sub-domain of physics, and we argue that we should be attending to this domain specificity." (Dreyfus *et al.*, 2019)

The research suggests that students make spontaneous epistemological comparisons between the two fields and shows that both *before and after instruction in quantum physics students report different epistemological beliefs* about classical physics and quantum physics.

Bifurcated surveys can be used to study domains below the school science subjects, and are also used to test *whether students' perception of expert views is aligned with their opinion*. In these cases, different perspectives are usually used, one stressing the expert position (the answer a scientist/physicist would give) and one stressing the personal position (what you really believe, or the answer that best expresses your feeling) (Gray *et al.*, 2008). In general, large differences between the perspectives suggest *disaligned views*, supporting the assumption that undergraduate students are willing to differentiate their perspective and express their own views as differing from the "right answer." Various teaching modes can be studied by differences in alignment, connecting the uptake of content with changes in epistemological beliefs. If post-instruction data show *good alignment*, it suggests that the internalization of the expert perspective is successful.

Whether bifurcated surveys are built on content surveys (the force concept inventory FCI) or epistemological surveys (the CLASS or MPEX), the modifications can be made to increase domain-specificity (along content lines, Maths vs Physics, classical vs quantum physics) or along epistemological lines (expert perspective vs individual belief). Dreyfus *et al.* (2019) provides a useful general discussion, and conclude that studying the sub-domain specificity of students' epistemic beliefs "can be fruitful and ought to influence our instructional choices." (*ibid.*)

Among other issues, gender *differences* can also be studied, suggesting that some earlier research (e.g., Perry's) might have been led astray by his choice of—mostly male—respondents (Kessels, 2013). Using bifurcated surveys, Adams *et al.* (2006) found that male perspectives about physics and learning physics are more aligned, while for female respondents the expert perspective is more disaligned with individual belief. Some research explicitly focuses on the gender-specific prevalence of underachievers in physics; for example, Hofer and Stern determined *gender-specific student profiles* using latent profile analysis (2016).

Focusing on relevant aspects of specific domains has triggered much innovative research. Conceptual knowledge in physics significantly changes through the course of one year, and so can epistemic beliefs

(Bigozzi *et al.*, 2018). There is growing evidence that epistemic beliefs vary not only as a function of academic domains but also that the variation is related to domain structuredness (e.g., between mathematics as a well-structured domain and education as an ill-structured domain). Below, I discuss only one nontrivial and interesting feature of this kind of research. *Advanced epistemic beliefs such as viewing scientific knowledge as uncertain can travel with unwanted features such as lower self-efficacy toward learning science*, as found when investigating the relationships between scientific epistemic beliefs, conceptions of learning science, and self-efficacy of learning science (377 Taiwanese high school students, Tsai *et al.*, 2011).[2]

At first sight, counterintuitive results show that certain "*expert-like*" *epistemic beliefs are not beneficial for learning in some cases*. These apparent anomalies are not only present in physics education or science education but also in other non-homogeneous domains. In fields where various types of abstractions are present, topics or domains that are well-structured or have narrow rules can provide exceptions. A case in point is vocational education and the training of business professionals, where epistemic beliefs were investigated on three tiers: general, domain-specific, and topic-specific level:

> For example, the more apprentices believe that knowledge in business administration is relevant for finding solutions in professional situations, the poorer their grades are in accounting (domain-specific level). At the same time, apprentices achieve better grades in accounting the more they believe that knowledge in accounting is relevant for professional situations (topic-specific level). … In the case of motivation in accounting, believing in complex business administration knowledge (domain-specific level) decreases motivation, while believing in complex accounting knowledge (topic-specific level) increases motivation. The reason for this may be that learners feel overwhelmed by the complexity of an entire company and the need to represent this company complexity within accounting. In contrast, experiencing only accounting knowledge as complex may increase the motivation to learn accounting because learners realize how the different accounting rules and systems work together and influence each other. (Berding *et al.*, 2017, p. 111)

This example from another area is detailed to show that domains where strict, narrow rules are essential (like accounting or mathematics and physics in STEM education) can behave unlike other domains. Domain-specific research is generally on the rise, with the danger of researchers focusing narrowly on one field or only on two tiers of abstraction (general and domain-specific beliefs). Domain-specific epistemic beliefs can have a more predictive power than general epistemic beliefs, and even more specific (topic-specific) beliefs can have an even better correlation with high motivation and

---

[2] Other results are more along the expected lines: "The analysis of the structural equation model revealed that students' absolutist scientific epistemic beliefs led to lower-level conceptions of learning science (i.e., learning science as memorizing, preparing for tests, calculating, and practicing), while sophisticated scientific epistemic beliefs might trigger higher-level conceptions of learning science (i.e., learning science as increase of knowledge, applying, and attaining understanding). The students' lower-level conceptions of learning science were also found to be negatively associated with their self-efficacy of learning science, while the higher-level conceptions of learning science fostered students' self-efficacy." (Tsai *et al.*, 2011).

achievement. To improve our understanding of the relationship between performance in learning areas and specific beliefs, it is just as crucial to study domain-specific epistemological beliefs in more detail as it is to incorporate results from other domains, acknowledging that a comparative and interdisciplinary approach can help coordinate and optimize both the research tools, and the teaching strategies.

11.2.6 Analysis of components: The emics of epistemic beliefs

In anthropology and sociology of science, a long debate has been linked to the differences between etic and emic approaches. Etic approaches usually allowed for better comparability (taking an outsider perspective with fixed reference), while emic approaches were "following their targets," and adapted the language of analysis to the local research. In this sense, while early approaches were "etic" frameworks, like a domain-general survey of EB, many of the current approaches are "emic."

The emics of epistemic beliefs provided *an ever-finer grain size of analysis* in the course of the ongoing debate on how to model a person's epistemology. Are there smaller "constituents" to epistemic beliefs that are productive resources for the construction of more sophisticated beliefs, e.g., the move from "knowledge is certain" to "knowledge is contingent on context and perspective"? To obtain a description of conceptions considered fine-grained enough, various terminologies were proposed (like "Resources" introduced by Elby and Hammer, 2001). Some researchers focused more on the context, the learning situation, or on what students perceive as relevant, but some concentrated on the cognition of the agent, how phenomenological primitives contribute to students' learning of physics (for links to epistemology, see Hammer, 1994; Hofer and Pintrich, 1997; and Hofer and Pintrich, 2002).

Modeling "high resolution" conceptual change tackles a fundamental theoretical difficulty in accounting for the emergence of sophisticated understanding (an influential source for model-based reasoning is Nersessian, 2008). This strand of research is itself heterogeneous; Domert *et al.* (2007) note "Apart from the magnification scale, these models also differ as to the form of epistemology, whether it is explicit or implicit for the student, and how context-dependent it is."

As the focus shifted from a longitudinal analysis of large cohorts, some articles tackle only *singular learning events* of single students. Let us take one model example. To account for *epistemologically relevant building blocks*, diSessa (1993) introduced p-prims. These primitives do not exist in isolation but are context-dependent and are coordinated in so-called coordination classes (diSessa and Sherin, 1998). In a study of a particular learning path to understand Newton's law of cooling and exponential decay, the p-prims are considered to be similar to physical laws in the sense that they prescribe what happens in situations to which they apply

> "They are 'what just happens, naturally.' However, there are many more p-prims than principles of physics, and, as knowledge elements, p-prims have rather different qualities compared to principles of physics. They are 'subconceptual' in the sense that they are not, in themselves comparable in complexity to scientific concepts, principles, or theories. However, p-prims do

> become part of the encoding of normative physical concepts and laws. P-prims are only weakly linked to language (there is no conventional lexicon for them), and a lot of their properties flow from their contextuality, exactly when they are invoked or not invoked." (diSessa, 2017)

Di Sessa stresses p-prims' malleability, as they attach to the various features of the world and can provide common-sense (and correct) results when bound to certain features of the world, but also "misconceptions" when bound to others. In this view, p-prims do not get replaced with other structures during the development toward expert understanding; instead, specific p-prims are activated in specific situations for people with more expertise.

In the study of personal epistemology and epistemic cognition, we can detect a shift in the use (and possibly meaning) of "epistemology" and "epistemic." Epistemic cognition for the Reflective Judgment Model was a high-level function of a rational agent that monitored the epistemic nature of problems. In contrast, when looking at building blocks di Sessa talks little about the agency of the learner and considers it a category error to construe p-prims to be true or false: "Ecologically, they work well enough in the circumstances in which they are normally evoked" (*ibid.*). It appears that knowledge-elements are context specifically activated; however, if the primitives are activated by the context and the learning situation (like p-prims), how can they contribute to Level 3 functioning (reflective judgment)? In this book, the chapter discussing "conceptual change" by Levrini investigates the tradition in detail. Below are two examples that focus on the conceptual development of students and are related to teaching abstractions, symbols and formalizations.

The study of epistemic beliefs can be linked to *teaching important abstractions*, which in turn can help improve the teaching materials. Mickey *et al.* (2017) studied the role of epistemic beliefs in gaining an understanding of the unit circle. Instead of focusing on various rules that apply only to a limited set of problem types, the authors focused on a broader conceptual framework around the unit circle "that supports an integrated problem solving procedure that can be applied across a range of problem types." *For this type of instruction, seeing mathematics as an integrated system of relationships is the preferred epistemic belief*, but other mindsets are also conducive to acquiring an integrated conceptual representation:

> … Students with a growth mindset also more strongly endorsed the idea that effort is necessary to learn, which may play a role in expectations about the quickness of learning and thus the ability of students to be patient and take the necessary time to learn, rather than rushing through materials for the sake of getting through them. Because these factors seem likely to us to have important influences on a student's ability to master the unit circle as an integrated conceptual structure underlying trigonometry, the next version of our materials will explicitly encourage viewing trigonometry as an integrated system of knowledge that requires time and engagement to learn. The materials will be introduced as an integrated framework as the students start into the program of lessons, and each block of the lesson will be described as playing a specific role within an integrative approach to understanding the meaning of trigonometric expressions within the unit circle framework. (Mickey, 2017)

Table 11.5
Epistemological components to map university physics students' epistemological mindsets towards the understanding of physics equations (analyzing interviews, Domert *et al.*, 2007, p. 20, abbreviated).

Epistemological component	The appropriate disciplinary epistemological mindset
A	Understanding involves being able to recognize **the symbols in the equation** in terms of the corresponding physics quantities
B	Understanding an equation involves being able to **recognize the underlying physics of the equation**
C	Understanding involves recognizing **the structure of the equation** —understanding how the different quantities in the equation are related to each other and the equation as a whole in terms of where the quantities are situated in the equation and what this infers.
D	Understanding involves **establishing a link between the equation and everyday life** —situating the equation in an everyday context by identifying examples and situations in everyday life where the equation applies —finding analogies from everyday life that help in appreciating the meaning of the equation
E	Understanding involves knowing **how to use the equation** to solve physics problems —using the mathematical manipulations that are needed to extract the sought information from the equation. —identifying which information is sought as well as what other information is available or needed.
F	Understanding involves being able to know **when to use the equation** —knowledge of the range of validity of the equation, inherent approximations and idealizations and in some cases also what branch of physics the equation is supposed to describe

Scaling up from a single integrated conceptual representation, one can study *classes of representations*. Table 11.5 shows an attempt to develop the components of physics students' epistemological stances when it comes to *the understanding of physics equations* (based on Domert *et al.*, 2007, abbreviated).

As such, the example in Table 11.5 while specific for physics relates to several of the dimensions of multi-dimensional research. It is easily operationalized and amenable to a "checklist" approach to help compare various aspects of understanding a physics equation; however, the checklist has no clear general epistemological implications and does not differentiate various stages of complexity, although it could be argued that most dimensions are more elaborated by experts.

11.3 FOUNDATIONAL ISSUES

The previous section provided a partly historical and partly conceptual narrative of the development. This section gives an overview of some issues that prospective researchers or users of established measuring tools should pay attention to before interpreting the results of a study.

Theoretically mapping the maturation of a growing and increasingly complex system is clearly underdetermined and thus various frameworks can provide some mapping of cognitive development. To meaningfully discuss personal epistemologies, our current assumption is that the knower has a gradually developing web of beliefs, where as the years pass, moving from accepting authority, the knower develops a refined and reasonable worldview, one that is responsive to criticism yet resilient.

Modeling a basic *explanandum* remains an issue: As students learn, the complexity of the outcomes of learning increases, and this is connected to the increasing structural complexity during development. Laying the appropriate conceptual foundations for the emergence and measurement of epistemic beliefs necessarily has loops of dependence between the proposed theory and evidence. A complex dynamic systems perspective, for example, implies that evolution is often nonlinear, dynamic conceptual structures are embedded in other dynamics, and new dynamic conceptual structures evolve from existing ones via periods of perceived disequilibrium (Brown, 2013).

The introduced tools exploring the problem of beliefs on knowledge and learning at the research frontier had specific purposes and contexts of development, but before a paradigm evolves for a field, there are various alternative conceptions on the right methods, approaches, and norms for what counts as proper research. Research in the last decades has been increasingly linked to educational agendas, school science subjects and grades instead of the general developmental issues of brain and cognition. For example, "domain" usually referred to the psychological domain for the earlier developmental tradition (executive, language learning, quantitative sense), but for current educational research it refers to school science subjects (physics, biology, and mathematics).

The section turns first to conceptual issues and the problem of the experimenters regress before a survey of some self-report questionnaire-items.

11.3.1 Conceptual issues

The salient variation with respect to conceptual issues is at the heart of the problem of finding an adequate theory of physics-specific epistemic beliefs. The blurred margin of the concept of "epistemic belief" has import on the most fundamental questions, such as *what* is measured. For historical reasons, models were developed with a focus on "lay" epistemologies; therefore, their applicability to measure expert-like epistemic beliefs is non-trivial to start with.

Furthermore, the term "epistemic" is inherently nebulous. Even in mainstream epistemological literature, concerns have been raised due to the *ambiguity in the use of "epistemic"* as it is used to mean *"of or relating to knowledge"* and also to mean *"of or relating to belief."* Hazlett even considered "the prospects for eliminating 'epistemic' from our philosophical lexicon" (Hazlett, 2016). Defining knowledge as "justified true belief" implies that reasons/justification are fundamental for epistemological development. Even if we prefer other definitions, for both laymen and contributory experts in specific fields, *some* mature epistemic beliefs should contribute to knowledge. While these beliefs should be considered as *epistemic* (linked to everyday knowledge as well as expert knowledge), the research on epistemic beliefs in general does not probe the *basing relation* “The epistemic basing relation is the relation which holds between a reason and belief if and only if the reason is a reason for which the belief is held. It is generally thought to be a necessary but not sufficient condition for belief to be justified that the belief be based on a reason.” (Korcz, 2021). When epistemologists distinguish good reasons that contribute to the personal justification of a given belief from good reasons which the person possesses but that do not contribute to the personal justification of the belief, they realize that the latter are not used as reasons for a belief. Responses in interview situations for the Reflective Judgement Model can in principle map the inferential and reasoning aspect and ill-structured problem-solving and can possibly be related to “knowledge.” In self-report questionnaires, however, responses are not differentiated according to the two categories, presumably measuring the latter category. Much of the multi-dimensional research collapses the distinction between having justified beliefs and having beliefs and maps not the justifications for (epistemic) beliefs, at most beliefs about justification.

A person holding a proposition about knowledge and knowing as true has a mental state, and “personal epistemology” assumes some coherent, reason-based belief-system. Although “content” knowledge is usually referred to as knowledge, in much of EB research “beliefs” are measured, akin to views on Nature of Science. Epistemic beliefs, it seems, do not constitute “knowledge”. To better understand epistemic beliefs, we can try to pin down “belief,” but in the research field notable heterogeneities can be found. They can be studied both as cognitive beliefs that have epistemic significance when students acquire content-specific target knowledge and as meta-level beliefs (e.g., conceptions on Nature of Science). Kitchener discusses epistemic beliefs as in the abstract sense not referring to knowledge (and knowing) but as specific instantiations of knowledge and knowing. Hofer and Pintrich (1997) note the mixed use in the research tradition: they either refer to cognitive structures (mental states and beliefs) or epistemic beliefs are defined as of through which an individual comes to know in specific contexts (cognitive processes).

Although social negotiation between educational researchers in the field resulted in some convergence, phenomena derived from research data may vary on the proposed framework of student epistemologies. Interpretation is linked to endorsing some (presumably apt) approach to “read” the result, and for one, a certain data-point can have various meanings. Table 11.6 gives some alternatives

Table 11.6

A response to a questionnaire item can be interpreted in various ways: What is measured affects the possible interpretations of the results. Attitude research is common in social psychology, but the three tiers are rarely used in research on EBs.

	Actions	Beliefs	Judgement	Attitudes
Level of description	Behavioral (voluntary)	Self-reflecting, Intentional (honest)	Rational (honest)	3-tiers: Cognitive, Affective, Conative
Propositional	?(explicit/implicit)	+	+	(+) via operationalization
Reasoned/Argumentative	–	?	+	–

to the interpretation of a questionnaire-item. Students' *actions* (picking response-options) are often interpreted as *epistemic judgments*, decisions about particular knowledge claims in specific contexts, and researchers can infer to *epistemic beliefs*, assuming that judgments of questionnaire-items are related to an individual's abstract beliefs about knowledge and knowing. One, however, need not assume abstract beliefs and can instead think of epistemic beliefs as parts of strategies utilized when individuals engage in epistemic cognition, with potentially no structure and highly context-dependent activation from an array of resources.

Bearing in mind that educational use is rather muddled (for some epistemological issues, see Kitchener, 2011), it is not surprising that the research on the topic is split along a number of "fault lines," incorporating various types of differences of opinion. Based on Sandoval *et al.* (2016), these relate to

Conceptualizations of Knowledge and Cognition—How individual and social aspects of epistemology are handled, but also which educational approach is used (e.g., constructivism, sociocultural theories of cognitive development, situated cognition). Some of the approaches might be incommensurable to one another.

Generality and Context—The question of domain generality and specificity remains an issue, as already discussed. Knowing that variation in response to situational demands are obviously large (as in the contextualist resource framework of Hammer and Elby), how can we study both development *and* context? While there is much variability across the epistemic norms and practices of disciplines, the traditional cognition components (dimensions) are still assumed to be applicable across disciplines and situations.

Developmental Pathways—Sandoval *et al.* notes: "what appears to be broad developmental trend may be an artefact of researchers' own assumptions and efforts to document that trend… Much more conceptual and empirical work is needed to distinguish models and mechanisms of epistemic change from epistemic development" (Sandoval *et al.*, 2016)

Research Design and Methods—Psychometric concerns can be raised for most standard tools. Various issues have been raised concerning the reliance on self-report response formats, for example, using Likert-type response scales.

Some of these issues are discussed in the next subsections.

11.3.2 Experimenter's regress

Many epistemologically relevant questions are built into the conceptual and methodological framework of an experiment. The *designed length of studies*, for example, encodes important assumptions. Long-term studies usually see development well into adulthood, and little salient change occurs in the course of 1 or 2 years. The Epistemological Reflection Model (by Baxter Magolda in 1992) used eight years of postcollege stories to link education with adult life: young adults were still significantly changing their epistemic beliefs. With the growing popularity of questionnaires, much current research is considerably shorter-term, focusing on a semester or a few years and arguing for measurable improvement. Short-term studies where *functional/optimal performance* is not differentiated provide little support—even when larger population sizes are employed[3]—whether the pre- and post-instruction results show a shift in the typical modes of functioning. Is it possible that many of the short-term findings can be considered "noise" for long-term research, artefacts produced by differences between the conditions for pre- and post-test, the earlier corresponding to functional performance, and the latter closer to optimal performance due to more context during instruction? Which perspective is valid? Should a good research instrument measure no significant change over a semester (like CLASS and MPEX results in many cases)?

With the introduction of any of the previously mentioned research designs, *a successful outcome to a study needs to be determined without an earlier existing consensus*, so the interpretation of the result—and therefore the outcome of research—is affected by the so-called *experimenter's regress*. Harry Collins coined this phrase when studying the purported credibility of claims for the existence of high fluxes of gravitational radiation. How can you build and test a reliable gravity-wave detector before even knowing for certain that gravity waves exist? Can you prove their existence with a gravity-wave detector that is not validated and calibrated? The sociologist notes what also applies to EB research: "where the detection of a novel phenomenon is in question, it is not clear what should count as a 'successful outcome'—detection or non-detection of the phenomenon" (Collins, 1981, p. 34).

Epistemological beliefs were first assumed to exist, and various approaches were developed to measure changes in EBs. Before the existence of validated surveys, *dogmatic foundations* were used to help in the development of early models. These models generally assumed broadly linear development, such as increasing complexity or the maturation of an ideal, critical, rational thinker. In the first case, a truism of cognitive development translates uneasily to possible epistemological stances, and in the second the epistemological desiderata, that is, the end goals of development are not easily operationalizable. When measuring beliefs, the assumption of a linear development and a "naïve"–"sophisticated" continuum became widespread. For the last few decades, epistemic beliefs have been studied mostly in the context of personal epistemologies and the various dimensions are often correlated with achievement (e.g., grades and test scores)—also some linear scale.

Widespread *factor analysis (salient from the 1990s)* became instrumental in standardizing research methodologies, and also to develop various alternatives. The typical research questionnaire, usually established based on extensive exploratory work, assumes that there is a particular construct that maps all epistemologies. Once the theoretical rationale for epistemological belief factors was problematized, repeated attempts were made to refashion the concept (and dimensionality) of "epistemic belief" or to

---

[3] Exploratory factor analysis (EFA) recommendations regarding the appropriate sample size have become more detailed, and conditions are better known in which EFA can yield good quality results for N below 50 (Mundform *et al.*, 2005), de Winter *et al.* (2009) gives estimates for the minimum required N for different levels of loadings (λ), number of factors (f), and number of variables (p). That is, approaches could be developed for classroom-size research, but early research and current trends are usually less pragmatic.

streamline an already proposed construct. Together with both qualitative and quantitative research, the interpretation of results produced novel phenomena, but the attempted stabilization of measuring epistemic belief has been problematic, similar to a new phenomenon in physics. Beliefs are directly not observable, and it is unclear how to look for those that have epistemological relevance. Although the research tradition has now developed for two generations, foundational issues of the metrics are still unresolved.

If EBs are structured, what is the best way to represent them? Even within one strand of research (multi-dimensional view of attitudes and beliefs), there are various conceptualizations. Epistemological understanding is elusive partly because of the various progressive tools: Views About Sciences Survey (VASS) has six conceptual dimensions, Schommer's research hypothesized five distinct dimensions of epistemology, often streamlined to a four-dimensional model of epistemic beliefs in science (Conley *et al.*, 2004). For Schommer, these beliefs were potentially disconnected, while some other researchers assumed that epistemic beliefs of individuals are coherent and integrated, akin to a theory (Hofer and Pintrich, 1997)

> "If, however, beliefs are not picked with reference to their potential to map to successful learning and comprehension, but as a theory of knowledge, then beliefs about learning (quick learning) and intelligence (fixed ability) can be disposed of. Hofer and Pintrich (1997) focused on the nature of knowledge (certain knowledge and simple knowledge) and knowing (omniscient authority), and 'purified the construct.'" (Bråten, 2010, p. 212)

Once we find that there is more support for certain dimensions (Simple Knowledge and Certain Knowledge) than others (Fixed/Innate Ability) should we drop some dimensions? If beliefs about quick learning can link to performance or mastery goals and can have motivational power, then should beliefs about "quick learning" be considered epistemic? Should beliefs about "nature of knowledge" be studied as a perspective that includes implicit theories of intelligence and ability or not?

It can just as easily be debated what a dimension regarding the nature of knowledge is as what is not. If the research is directed to help students learn how to learn, the problem of *potential confounding factors* and other hypothesized factors becomes prominent. When developing the Reflective Judgment Model, reflective judgment scores were assumed to be affected by confounders**.** Not just competing theoretical constructs were investigated (Piagetian formal operations, chemicals, and pendulum tasks) but also verbal ability (using Terman's Concept Mastery Test) as well as verbal fluency, checking the number of words used in the interview (Kitchener and King, 1981). For the Reflective Judgment Model, reflective *judgment level was found to be closely related to verbal ability*. While the other four factors (including socio-economic status) seemed less important, verbal ability was a good predictor of reflective judgment level.

With quick cross-fertilization of fields, the number of potential confounders quickly increases. Not taking into account the scores from many possible tests can result in overestimating the impact of epistemic beliefs. Rudloff *et al.* (2022), for example, found that individuals who show a general tendency to maximize their personal utility are more likely to hold post-truth epistemic beliefs, such as high faith in one's intuition for facts, comparably low need for evidence, and viewing truth as being shaped by those in power. Interestingly, a scale measuring aversive personality and selfish, unscrupulous behavior can be a confounder for epistemic beliefs. Can the Dark Factor of Personality (high D-scores, for test-development, see Moshagen *et al.*, 2020) be a potentially good predictor of a certain set of epistemic beliefs? Apart from other *durable* traits (political orientation, personality-constructs), potential confounders can include various traits assumed to be *less stable* than beliefs like epistemic emotions, motivation (that can be extrinsic or intrinsic). The ontological assumptions behind the interpretation of questionnaires show great variation, as similar self-reporting techniques (e.g., Likert scales) can just as easily measure *orientation* as *belief, personality*, or even *emotional state.* Although many scientific articles suggest causal relationships where EBs are shown to have an effect on, e.g., achievement, the ontological flexibility of interpreting self-report questionnaires makes it challenging to identify the theoretically reasonable links, to correctly characterize cause-and-effect relationships, and to convert results to classroom use.

Moving from debates concerning the dimensions of research to subsets of questions and individual items of questionnaires, the next section discusses the ambiguity concerning interpretation and the way "epistemic beliefs" are abstracted from responses to self-report items.

11.3.3 The adequacy of the measurement-tools

Epistemic beliefs are often inferred from actions such as filling out questionnaires. How well can graded agreements with a statement-item be used to measure sophisticated epistemic beliefs? Measurement of a feature or aspect of EBs is nontrivial; for example, the quantitative methodologies that use validated sets of items do not inform us whether a self-assessed judgment is justified (supported by reason) or not. Such methods are agnostic about what counts as an appropriate meta-belief, i.e., a reason that is a good reason to hold a specific epistemic belief.

There are good research incentives to use existing tools and gather larger data-sets, but there are also valid research incentives to optimize the tools, test their applicability, and refine them. Likert scales, the common measurement scale for epistemic beliefs carry no justificatory connotations in their traditional interpretation and it has been found early on that the items are open to a wide variety of interpretations by both respondents and researchers. Schommer's items, when studied not in subsets but individually, did not always factor on the subsets they were originally designated to.

Although Likert-scales admittedly have limitations, nevertheless, they have been widely used for self-assessment of epistemological beliefs. Two respondents may express opposite positions on a Likert item for the same reason or the same position for contradictory reasons (Aikenhead, 1988):

> "Regarding ambiguity in student assessment, the Likert-type responses were the most inaccurate, offering only a guess at student beliefs. Such guesswork calls into question the use of Likert-type standardized tests that claim to assess student views about science. Student paragraph responses

> contained significant ambiguities in about 50% of the cases. The empirically developed multiple choices, however, reduced the ambiguity to the 20% level. Predictably, the semistructured interview was the least ambiguous of all four response modes, but it required the most time to administer. These findings encourage researchers to develop instruments grounded in the empirical data of student viewpoints, rather than relying solely on instruments structured by the philosophical stances of science educators." (Aikenhead, 1988)

The tool is not optimal for interpreting the participants' tentative positions, and measurement of naïve belief for certain individuals might be an artefact. In cases when interpreting the low and high scores is unproblematic, the scores in between can provide challenges. Early scales were normative, using statements with possible answers that covered a range of answers from "good" to "bad," but some deviated from this assumption. Among several attempts to refine this type of questionnaire, some maintain the scale but *weaken the normative suggestion usually linked to the scale*. For example, the Views About Science Survey (VASS) Implementation Guide by Adrian Madsen and Sam McKagan uses a contrasting alternative design for descriptive purposes:

> "…where students are given two viewpoints and asked to compare and contrast two things, whereas the CLASS and MPEX use a standard 5-point Likert scale. The expert-like response to the questions is not always clear, whereas the expert-like response on the CLASS and MPEX is clear. The VASS is useful for discussing the ideas around students' beliefs about learning physics, but is less useful for reliably measuring how expert-like your students' beliefs are." (Halloun and Hestenes, 1998)

When publishing research in journals, the concise format leaves little room to discuss the various self-report questionnaire items, yet questions can be introduced, discarded, multiplied, and modified. Table 11.7 lists and critically probes some of the items in established surveys to show differences in grading (positions that are considered "naïve" and "expert" views) and also some possible critical questions of interpretation.

11.3.4 From methodological questions to the perceived congruence of the field

This section on foundational issues deals with the problems of interpreting results, related to the question of the perceived congruence of the field. One characteristic of an emerging field is that while some take the existing tools to be *technical things with stable and reliable properties*, others still treat the tools themselves as *epistemic objects or things, not completely understood and not reliably established* (Rheinberger, 1997; and Chang, 2011). Let us take an example. In some meta-analyses, the distinctions between research tools are collapsed, and various surveys (and scales) are displayed as commensurable: "The CLASS and MPEX are similar in the way they measure students' beliefs about physics and learning physics, so the scores for these tests have been combined" (Madsen *et al.*, 2015). The tool is used as a standard tool, e.g., thermometer, and research gathers and builds on the data. In contrast, systematic integration of evidence can be carried out much more cautiously, not assuming

Table 11.7

Questionnaire items with critical questions (*For VASS items 21 and 24 note by Halloun and Hestenes, 1996).

Survey type and item number What and how is measured	Test survey item and notes on measurement	Epistemological issues and Critical questions
EBAPS 25 Source of ability to learn (a) I agree almost entirely with Anna. (b) Although I agree more with Anna, I think Emily makes some good points. (c) I agree (or disagree) equally with Anna and Emily. (d) Although I agree more with Emily, I think Anna makes some good points. (e) I agree almost entirely with Emily. A = 0, B = 1, C = 2, D = 4, E = 4	Anna: I just read about Kay Kinoshita, the physicist. She sounds naturally brilliant. *Emily: Maybe she is. But when it comes to being good at science, hard work is more important than “natural ability.” I bet Dr. Kinoshita does well because she has worked really hard.* Anna: Well, maybe she did. But let’s face it, some people are just smarter at science than other people. Without natural ability, hard work will not get you anywhere in science!	Example grades *productive* stance (based on empirical educational research), not necessarily a *realistic* stance for all students. Answer (e) is more productive, answer (d) is more realistic, and—unlike for item 28—both answers yield 4 points.
EBAPS item 28 Evolving knowledge (a) I agree almost entirely with Leticia. (b) I agree more with Leticia, but I think Nisha makes some good points. (c) I agree (or disagree) equally with Nisha and Leticia. (d) I agree more with Nisha, but I think Leticia makes some good points. (e) I agree almost entirely with Nisha. A = 0, B = 1, C = 2, D = 3, E = 4	Leticia: Some scientists think the dinosaurs died out because of volcanic eruptions, and others think they died out because an asteroid hit the Earth. Why can’t the scientists agree? *Nisha: Maybe the evidence supports both theories. There is often more than one way to interpret the facts. So we have to figure out what the facts mean.* Leticia: I am not so sure. In stuff like personal relationships or poetry, things can be ambiguous. But in science, the facts speak for themselves.	Is the question outdated (for the particular example)? Is “in science, the facts speak for themselves” a bad answer generally or in all cases? If (E), a moderate / tentative position is preferred (evidence supports both theories), why isn’t (D) the most sophisticated answer? (problem of reflexivity)

MPEX item 19 Context dependence Favorable: believes physics needs to be considered as a connected, consistent framework Unfavorable: believes physics can be treated as separated facts or "pieces" 5-scale Likert, agree-disagree	*The most crucial thing in solving a physics problem is finding the right equation to use.* Unfavorable response "A" indicates agree or strongly agree—a choice of numbers 4 or 5. Favorable response "D" indicates disagree or strongly disagree—a choice of numbers 1 or 2.	Is the question meaningful (enough) without context? Does negative response (expert view) *entail* belief in physics as a connected, consistent domain? No distinction near the expert end of the spectrum (D is both response 1 and 2)
Views About Science Survey (VASS) items 21 and 24 N1 + Science is about generic (a) coherently interrelated conceptions, and (b) patterns of thinking, including problem solving. – rather than about idiosyncratic and isolated, situation-specific terms, statements and procedures:	21. The laws of physics are (a) inherent in the nature of things and independent of how humans think. (b) invented by physicists to organize their knowledge about the natural world. 24. Newton's laws of motion: (a) will always be used as they are by physicists. (b) will eventually be replaced by other laws.	* The following two items were somewhat controversial, especially among university professors: In item 21, 50% were more inclined towards alternative (a), and 31% towards (b). The rest equally favored both alternatives. We are more in favor of (b), which we consider in this paper to be the expert view. In item 24, 58% were more inclined towards alternative (a), and 33% towards (b). We happen to disagree on this particular item. One of us (DID feels strongly in favor of (a), the other (IH) feels more in favor of (b). We will rewrite this particular item in a way that would clearly keep (b) the expert view.

that a construct is fully understood or reliable. Schiefer *et al.* (2022) carried out latent factor analysis on multiple samples before arguing for the existence of epistemic belief profiles as well as correlations with external student variables.

This lack of congruence can be displayed by other sets of examples. Regular variable-centered approaches, such as multidimensional research, usually *focus on populations*; therefore, MPEX issued a Product Warning Label for the Questionnaire:

> "Note that individual items from this survey should not be used to evaluate individual students. On any single item, students may have atypical interpretations or special circumstances which make the 'non-expert' answer the best answer for that student. Furthermore, students often think that they function in one fashion and actually behave differently. For the diagnosis of the difficulties of individual students more detailed observation is required. This survey is primarily intended to evaluate the impact of one or more semesters of instruction on an overall class. It can be used to illuminate some of the student reactions to instruction of a class that are not observable using traditional evaluations. In this context, it, together with evaluations of student learning of content, can be used as a guide for improving instruction." (Redish, 2001)

For the last decades, *person-centered models* have been on the rise, yet it is not obvious whether the classical approaches designed to study populations (e.g., by exploratory factor analysis) should be replaced or used nevertheless.

As EBs are studied in a heterogeneous research tradition, many of the foundational questions are not settled. Students can simultaneously hold both domain-specific and more domain-general or overarching epistemic beliefs (Buehl and Alexander, 2001; and Bråten, 2010), so epistemic beliefs are not *only* domain specific, but interpretation of the relationship (and reliability) of domain-specific, and -general responses can vary. The legitimacy of the research constructs, for example the assumption in domain-specific CLASS that physics is a single epistemic domain, cannot be easily verified by the respective framework.

Some perplexing results may have to do with the inadequacy of the statements in the questionnaires, and others with foundational issues. Although epistemological beliefs in some cases predict academic achievement, in some cases "naïve," less sophisticated epistemic beliefs can yield better results. Does this invalidate a specific dimension of the construct? Or the dimension should be used, but the scale is not a naïve-sophisticated continuum? As the perceived congruence of the field is significantly different for the practitioners, convincing the community that there is a particular construct that maps all epistemologies will remain difficult.

11.4 ON SOME OF THE MANY ROADS AHEAD

Research on EBs is widespread, from high schools to teacher-training across the globe. While more general epistemic beliefs seem to be relatively stable, the practices and norms of the classroom significantly influence students' domain-specific EBs. Current research on domain-specificity increases attention to specific knowledge and manipulation of that reservoir.

In teacher-education the study of EBs is likely to increase, together with a growing demand to study the *interactions between EBs and performance* (both academic and teaching performance). Studying epistemic beliefs in action shows that a teacher's naive epistemological beliefs are clearly reflected in the teacher's teaching practices during lab activities, "However, a teacher's sophisticated epistemological beliefs are not always clearly connected to the practice. This seems to be related to the necessary negotiation among their epistemological beliefs, teaching contexts, and instructional goals." (Kang and Wallace, 2005). The observational data gained from responses to critical incidents can be used to link differences in teaching actions and differences in EBs. Caleon *et al.* (2018) found that many of the early-career teachers held beliefs about learning physics that were incongruent with their beliefs about teaching physics, and classroom practices of novice teachers on the topic of electricity were more aligned with their beliefs about learning physics than with their beliefs about teaching physics. Bae *et al.* (2022) studying generative environments and pedagogical practices found that to change habits of teaching, teachers benefit from regular long-term support (min. 18 months).

As epistemic beliefs are relevant as early as elementary school, EB profiles can be used to help adapt teaching content to improve the uptake of early science-related courses and better understand learning profiles. With a growing emphasis on latent profile analysis that can locate subpopulations, the emerging *student profiles* can be studied with corresponding differences in scientific inquiry competencies, motivational dispositions, and social background (Schiefer *et al.*, 2021). EBs can also be matched with achievement goals from early on (Winberg *et al.*, 2019).

It is likely that certain issues can be studied from childhood, for example, reliance on authority (Source of knowledge) is probably an elaborate dimension from early on. *The social orientation* in development is present early on, and authority and *natural pedagogy* play a crucial role in development (Csibra and Gergely, 2011). Children appear to actively *monitor the reliability of a speaker's* knowledge claims (Kushnir and Koenig, 2017), who can distinguish unreliable speakers and can reject testimony from a previously inaccurate speaker. A belief-dimension akin to "Source" or "Authority" is plausibly needed to explain that children's appraisal of a speaker can even trump perceptual access to support a claim.

The development of other fields—or science education in general—suggests some avenues that might increase in significance. Mercier and Sperber's views on the *development of reasoning ability* (2009 and 2011) link confirmation bias and justification of one's own position to persuasion. Attending to the details of reasoning will probably help us much better appreciate the finer details of holding beliefs and the willingness to revise them. Young children (ages 4–11 years old) can be more overconfident in their knowledge and are also more likely to revise their initial beliefs (Hagá and Olson, 2017). Studying the dissociation between the confidence with which beliefs are held and the revision of those beliefs across development calls into question lay theories (positing anticorrelation of overconfidence and belief-revision) and links questions of epistemic trust with metacognitive issues. The difficulty of interpreting the participants' responses suggests that justification for knowing is probably itself a complex dimension, and a single hierarchical continuum can misrepresent it, implying that an appropriate construct has to provide room for multiple means of justification (see Greene *et al.*, 2008).

Throughout the educational research landscape, new and pressing needs are recognized. In the age of social media, there is a growing demand to develop updated tools that relate to scientific and media literacy, and help students understand internal and external science communication (Höttecke and Allchin, 2020). Recent "post-truth" scenarios also pose new challenges and opportunities; certain epistemic beliefs may cause harm to the individual or a group, like Covid-19 conspiracy theories (Rudloff *et al.*, 2022). EB research can link certain beliefs with the prevalence of conspiracist inclinations: Confidence in the ability to intuitively recognize truth is a uniquely important predictor of conspiracist ideation as found by (Garrett and Weeks, 2017).

Research on epistemic beliefs is rooted in the period before *teaching about the nature of science* became widespread. One of the most relevant changes to science education is the gradual appearance and spread of curricular *content on epistemology, rationality, and argumentation*, that are not traditional disciplinary contents. The changing focus of science education prioritizes earlier neglected areas, such as socio-scientific issues (SSI), ethical issues, incorporating reflection on science and values (Koster and de Regt, 2020; and Garrecht *et al.*, 2022). The image of science as value free is being replaced, and responsible teachers, researchers and citizens need to learn to articulate and reflect upon their own values. The need for innovation shapes normative agendas (Dwyer *et al.*, 2014), and as a result, many courses are likely to provide their own explicated epistemologies, their own take on knowledge-construction and uptake. The challenge of *Transdisciplinarity* has already been recognized (Sandoval *et al.*, 2016), and at universities, courses on ethical issues, "heuristics and biases," theories of bounded rationality, or decision-making have gained popularity.

The problems about which "rational people reasonably disagree" include the problem of defining and measuring epistemic beliefs as well as finding a place for such research in the future. On the one hand, to prepare for the increasingly complex and rapidly changing work-environment 21st-century skills and transversal skills are gaining ground, shifting attention to *skills* from *beliefs*. On the other hand, if science classrooms promote epistemology of science both as content of instruction and as embedded in instructional methods, then epistemic *beliefs* might benefit in the future from epistemic *content knowledge*.

REFERENCES

Adams, W. K. *et al.*, Phys. Rev. Spec. Top.—Phys. Educ. Res. **2**(1), 010101 (2006).
Aikenhead, G. S., J. Res. Sci. Teach. **25**(8), 607–629 (1988).
Bae, Y. *et al.*, Instr. Sci. **50**, 143–167 (2022).
Berding, F. *et al.*, World J. Educ. **7**(3), 103–114 (2017).
Bigozzi, L. *et al.*, Front. Psychol. **9**, 2474 (2018).
Bråten, I., *International Encyclopedia of Education*, 3rd ed., edited by E. Baker *et al.* (Elsevier, 2010).
Brown, D. E., Sci. Educ. **23**(7), 1463–1483 (2013).

Buehl, M. M. and Alexander, P. A., Educ. Psychol. Rev. **13**(4), 385–418 (2001).
Caleon, I. S. *et al.*, Res. Sci. Educ. **48**, 117–149 (2018).
Chang, H., Erkenntnis **75**(3), 413–429 (2011).
Collins, H. M., Soc. Stud. Sci. **11**(1), 33–62 (1981).
Conley, A. M. *et al.*, Contemp. Educ. Psychol. **29**(2), 186–204 (2004).
Csibra, G. and Gergely, G., Philos. Trans. R. Soc. B **2011**(366), 1149–1157 (2011).
Dewey, J., *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process* (D.C. Heath & Co Publishers, Boston, MA, 1933).
de Winter, J. C. *et al.*, Multivariate Behav. Res. **44**(2), 147–181 (2009).
diSessa, A. A., Cognit. Instr. **10**, 105–225 (1993).
diSessa, A. A., Hum. Dev. **60**(1), 1–37 (2017).
diSessa, A. A. and Sherin, B. L., Int. J. Sci. Educ. **20**(10), 1155–1191 (1998).
Domert, D. *et al.*, NorDiNa **3**(1), 15–28 (2007).
Dreyfus, B. W. *et al.*, Int. J. STEM Educ. **6**, 31 (2019).
Dwyer, C. *et al.*, Think. Skills Creat. **12**, 43–52 (2014).
Elby, A., Am. J. Phys. **69**(S1), S54–S64 (2001).
Elby, A. and Hammer, D., Sci. Educ. **85**(5), 554–567 (2001).
Fischer, K. W., Psychol. Rev. **87**(6), 477–531 (1980).
Fischer, K. W., "Dynamic cycles of cognitive and brain development: Measuring growth in mind, brain, and education," in *The Educated Brain: Essays in Neuroeducation*, edited by A. M. Battro, K. W. Fischer, and P. J. Léna (Cambridge University Press, 2008), pp. 127–150.
Fischer, K. W. *et al.*, *Handbook of Developmental Psychology*, edited by J. Valsiner and K. Connolly (Sage, Thousand Oaks, CA, 2003), pp. 491–516.
Fry, H. *et al.*, *A Handbook for Teaching and Learning in Higher Education* (Routledge, 2008), pp. 26–44.
Garrecht, C. *et al.*, Sci. Educ. (published online 2022).
Garrett, R. K. and Weeks, B. E., PLoS ONE **12**(9), e0184733 (2017).
Gray, K. E. *et al.*, Phys. Rev. Spec. Top.—Phys. Educ. Res. **4**(2), 020106 (2008).
Greene, J. A. *et al.*, Educ. Psychol. **43**(3), 142–160 (2008).
Hagá, S. and Olson, K. R., Dev. Psychol. **53**(12), 2319–2332 (2017).
Halloun, I. and Hestenes, D., *Paper Presented at the Annual Meeting of the National Association for Research in Science Teaching*, St. Louis, MO, March 31–April 3, 1996 (ERIC Clearinghouse, Washington, D.C., 1996), p. 32.
Halloun, I. and Hestenes, D., Sci. Educ. **7**(6), 553–577 (1998).
Hammer, D., Cognit. Instr. **12**(2), 151–183 (1994).
Hazlett, A., Episteme **13**(4), 539–547 (2016).
Hofer, B. K., Contemp. Educ. Psychol. **25**(4), 378–405 (2000).
Hofer, B. K. and Pintrich, P. R., Rev. Educ. Res. **67**(1), 88–140 (1997).
Hofer, B. K. and Pintrich, P. R., *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing* (L. Erlbaum Associates, Mahwah, 2002).
Hofer, S. I. and Stern, E., Learn. Individual Diff. **51**, 119–131 (2016).
Höttecke, D. and Allchin, D., Sci. Educ. **104**(4), 641–666 (2020).
Kampourakis, K. and McCain, K., *Uncertainty: How It Makes Science Advance* (Oxford University Press, New York, NY, 2019).
Kang, N. H. and Wallace, C. S., Sci. Educ. **89**(1), 140–165 (2005).
Kessels, U., Learn. Individual Diff. **23**, 256–261 (2013).
King, P. M., New Dir. Teach. Learn. **2000**(82), 15–26 (2000).
King, P. M. and Kitchener, K. S., *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*, Jossey-Bass Higher and Adult Education Series and Jossey-Bass Social and Behavioral Science Series (Jossey-Bass, San Francisco, CA, 1994).
King, P., Kitchener, K. & Wood, P., "Research on the reflective judgment model," in *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*, edited by P. King and K. Kitchener (Jossey-Bass, San Francisco, 1994), pp. 24–202.
King, P. M. *et al.*, Hum. Dev. **26**(2), 106–116 (1983).
Kitchener, R. F., Sci. Educ. **2**, 137–148 (1993).
Kitchener, R. F., *Links Between Beliefs and Cognitive Flexibility* (Springer, Dordrecht, 2011), pp. 79–103.
Kitchener, K. S. and Fischer, K. W., *Developmental Perspectives on Teaching and Learning Thinking Skills. Contributions to Human Development*, edited by D. Kuhn (S. Karger, Basel, 1990), Vol. 21, pp. 48–62.
Kitchener, K. S. and King, P. M., J. Appl. Dev. Psychol. **2**(2), 89–116 (1981).
Kitchener, K. S. *et al.*, Dev. Psychol. **29**, 893–906 (1993).
Korcz, K. A., *The Epistemic Basing Relation* (Stanford University, 2021), first published Oct 31, 2002; substantive revision Feb 25, 2021.
Koster, E. and de Regt, H. W., Sci. Educ. **29**, 123–143 (2020).
Kushnir, T. and Koenig, M. A., Dev. Psychol. **53**(5), 826–835 (2017).
Madsen, A. *et al.*, Phys. Rev. ST Phys. Educ. Res. **11**, 010115 (2015).
Mercier, H. and Sperber, D., *In Two Minds: Dual Processes and Beyond*, edited by J. S. B. T. Evans and K. Frankish (Oxford University Press, 2009).

Mercier, H. and Sperber, D., Behav. Brain Sci. **34**(2), 57–74 (2011).
Mickey, K. W. and McClelland, J. L., *Acquisition of Complex Arithmetic Skills and Higher-Order Mathematics Concepts* (Academic Press, 2017), pp. 247–269.
Moshagen, M. *et al.*, Psychol. Assess. **32**, 182–196 (2020).
Mundfrom, D. J. *et al.*, Int. J. Testing **5**(2), 159–168 (2005).
Nersessian, N. J., *Creating Scientific Concepts* (MIT Press, 2008).
Perry, W. G., *Forms of Intellectual and Ethical Development in the College Years: A Scheme* (Holt, Rinehart and Winston, New York, 1970).
Ramsden, P., *Improving Learning: New Perspectives* (Kogan Page, London, 1988).
Redish, E. F., *Student Expectations in University Physics: Using the Maryland Physics Expectations Survey* (University of Maryland Physics Education Research Group, 2001).
Redish, E. F. *et al.*, Am. J. Phys. **66**(3), 212–224 (1998).
Rheinberger, H. J., *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube* (Stanford University Press, 1997).
Rose, S. P. and Fischer, K. W., Br. J. Dev. Psychol. **16**(Pt 1), 123–131 (1998).
Rudloff, J. P. *et al.*, J. Pers. **90**, 937–955 (2022).
Sandoval, W. A. *et al.*, Rev. Res. Educ. **40**(1), 457–496 (2016).
Schiefer, J. *et al.*, Learn. Individual Diff. **92**, 102059 (2021).
Schiefer, J. *et al.*, Educ. Psychol. Rev. **34**(3), 1541–1575 (2022).
Schommer, M., J. Educ. Psychol. **82**, 498–504 (1990).
Schommer, M., Educ. Psychol. Rev. **6**(4), 293–319 (1994).
Schommer, M. and Walker, K., J. Educ. Psychol. **87**, 424–432 (1995).
Schommer-Aikins, M. and Duell, O. K., Rev. Invest. Educ. **31**(2), 317–330 (2013).
Sperandeo-Mineo, R. M., *Proceedings of the ESERA 2011 Conference: Science Learning and Citizenship* [European Science Education Research Association (ESERA), 2012].
Tsai, C.-C. *et al.*, Learn. Instruct. **21**(6), 757–769 (2011).
Vygotsky, L. S., *Mind in Society: The Development of Higher Psychological Processes* (Harvard University Press, Cambridge, MA, 1978).
Wang, S. *et al.*, Sci. Educ. (published online 2022).
Winberg, T. M. *et al.*, Eur. J. Psychol. Educ. **34**, 295–315 (2019).

CHAPTER 12 PHILOSOPHY OF PHYSICS: ITS SIGNIFICANCE FOR TEACHING AND LEARNING

Roland M. Schulz and Calvin S. Kalman

Schulz, R. M. and Kalman, C. S., "Philosophy of physics: Its significance for teaching and learning," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 12-1–12-30.

> *A knowledge of the historic and philosophical background gives that kind of independence from prejudices of his generation from which most scientists are suffering. This independence created by philosophical insight is—in my opinion—the mark of distinction between a mere artisan or specialist and a real seeker after truth.*
>
> —Albert Einstein, Letter to Robert Thornton, 1944

12.1 INTRODUCTION

Physics and philosophy share a long interdependent conceptual history that has not usually been recognized or appreciated, in particular, among physicists and physics teachers. What is underappreciated by these two groups must necessarily remain so for students. The viewpoint that physics and philosophy have little to do with one another is widespread (Rovelli, 2018; and De Haro, 2020)—especially in educational circles—some physicists themselves insisting that philosophy is best avoided (Weinberg, 1992; and Hawking and Mlodinov, 2010). That this perception persists is established by a cursory examination of any secondary or freshman physics textbook, despite decades of arguments to the contrary by science educators, philosophers and physicists alike (Cushing, 1989, 1998; Lange, 2002; and Matthews, 2015). The primary purpose of this chapter is to finally put an end to this myth. It is claimed that philosophy is required to not only *do* physics but also to improve it. The philosophy of physics entails philosophizing about the concepts, theories and experiments of physics, which should be encouraged at all three educational levels—secondary, undergraduate, and graduate. Further, the launch of the academic sub-discipline *philosophy of physics* as a small but thriving research field today involves both physicists and philosophers working on deep-seated mathematical and conceptual puzzles.

https://doi.org/10.1063/9780735425514_012

Nonetheless, this mistaken belief has resulted in a type of one-sided physics education, some call it "technical pre-professional training" (TPT), that has become representative of secondary physics classes across countries and curricula, certainly common among first year undergraduate courses (whether calculus- or non-calculus-based). The focus resides on mastering concepts and laws by memorization, formula manipulation and numerical problem-solving, primarily through textbook and direct instruction (Arons, 1997). Textbooks look remarkably alike from nation to nation,[1] and all can agree that standardization has set in, and there are certainly reasons for our current condition: the PSSC reforms initiated after the 1957 "Sputnik shock" (Matthews, 2015), the "logic" of the discipline structure, with the momentum carried by the economics of textbook writing and publishing. The issue is compounded because physicists as textbook authors, while specialists in their research areas, are hardly specialists in the history and philosophy of their field. Yet this familiar "system," and by now common culture, has trained generations of physicists and physics teachers, and without question its curricular efficiency and structural logic speaks to its dominance and continued use. But that does not mean curricula and textbooks cannot be organized differently or worse, that there have not been any harmful pedagogical consequences because of the status quo: many have bemoaned the ongoing decline of students entering secondary and undergraduate courses; students exit first year courses with a greater anti-scientific mind than first entered; and the distorted history of physics in many textbooks. Less known is the submergence, if not outright disappearance of philosophy due to an overt *operationalist bias*.

This has provided generations with a false sense of the development of the discipline, including its methodology, the image of scientists and the scientific enterprise itself, especially at the upper levels. When the *historicity* of concepts with their philosophical matrix is ignored, along with the true impact of fundamental breaks with former orthodox theories (i.e., revolutions), then progress is seen through the cloudy lens of a smooth transition from theory to theory,[2] corresponding to the view that the growth of knowledge is sequential, straightforward and cumulative: the dynamism, the debates and mistaken concepts and theories are lost to instruction and curricula (e.g., *Bohr*: Pais, 1991; and *Einstein*: Ohanian, 2008). Here, there is no room for philosophical thinking, only technical mastery. The pedagogy assumes that one cannot obtain a better understanding of concepts through their genuine historical or philosophical background.[3] Hence, what students are exposed to are not the interpretive issues (and controversies) in the development of key terms (e.g., space, mass, force, energy, fields) and theories (e.g., classical mechanics to quantum theory) but a selective and sanitized version, though they graduate with the illusion of having learned it.[4] What remains is a "rhetoric of conclusions" confronting

---

[1] Often repeating the same problem exemplars (i.e., Atwood's device), and strictly organized according to the same topics embracing the phenomena under study (mechanics, waves, light, electrostatics, and so on).

[2] Typical presentations masking upheavals in thought: shift from Newtonian ("an approximation") to Einstein's gravity; old quantum theory to quantum mechanics (no obvious crisis); "linear" progression of atomic theory from the Greeks to Dalton to Rutherford in physics and chemistry textbooks (under the guise of "theory revisions").

[3] See Coelho (2013) on classical mechanics problems that flatly contradict this prevalent assumption.

[4] Typical examples from secondary and tertiary textbooks: Lavoisier is praised for overthrowing phlogiston, but the fact that he mistakenly held to caloric theory and was anti-atomism is ignored. Likewise, Carnot is praised for hypothesizing a steam engine (Carnot cycle), but he too held to caloric. Everywhere, Dalton is praised for re-introducing the atomic theory in 1808 (though fraught with errors), yet he linked it to caloric and opposed kinetic theory.

the learner, restricting reasoning at best to the "normal paradigm" of final form science (see Kuhn on science education, 1970, pp. 136–143).[5] The status quo certainly does foster mathematical flexibility and problem-solving proficiency, and quite a lot of knowledge about terms, equations and laws, but not creative or critical thinking. Behind the important issues of the economy of space in textbook pages and the economy of the time needed for instruction, both of which weigh heavily against possible philosophical exposition, is the widespread prejudice that history and philosophy are irrelevant for learning "real" physics. It is no wonder students come away with that impression, alongside physicists and physics teachers themselves having been inculcated in that culture. It is certainly not seen as a "human adventure."

12.2 A BRIEF REVIEW OF HISTORICAL INTERACTION BETWEEN PHYSICS AND PHILOSOPHY

Let us therefore briefly survey the historico-philosophical background to see if physicists and teachers can come to recognize the value of philosophy *for* physics (so that later the nature of philosophy *in* physics and *of* physics becomes clearer), in order for their instruction and curricula to "make room" for philosophy.

The intimate relationship between physics and philosophy runs through Western culture regarding the development of human reasoning about the physical world, extending as far back as Pre-Socratic thinkers in Ancient Greece.[6] The conceptual analysis of physical ideas such as matter, motion, causes, time, infinity, and void (including Zeno's paradoxes) followed Aristotle's *Physics* (1996). The birth of mathematical physics can be traced to Archimedes (laws of buoyancy and levers). Astronomy began its break from astrology with research on the physical modeling of the heavens by Plato's student Eudoxus and the complex mathematical cosmology of Ptolemy (Kuhn, 1957; and Clagett, 1963). In philosophy, Plato made major contributions to the nature and theory of knowledge (*epistemology*) with his definition of "justified true belief" contrasting "opinion" in *Theaetetus* (2003), while Aristotle surveyed questions about the nature of reason, being, causes, and idea-lism (*ontology*) in his *Metaphysics* (1998). Philosophical atomism survived in Epicureanism (Lucretius, 1951).

In the Enlightenment Age (1650s onwards), there occurred a mutual and intensive interaction that helped to advance both fields, which itself represented an upheaval in Western thought, science and society (Kuhn, 1957): new discoveries, theories and investigative methodologies (of Bacon, Kepler, Galileo, Descartes, Huygens, Boyle, Newton, Leibniz), were buoyed by an emerging *mechanistic philosophy*; the reaction in philosophy took new directions with empiricism, rationalism, idealism

---

[5] The textbook plays a fundamentally conservative role since it reinforces the latest dominant paradigm and presents subject/content knowledge as "linear and cumulative" (p. 139). He admits this text-centered pedagogy stifles imagination and innovation. The Holton and Brush (2001) textbook aims at historical accuracy.

[6] Starting with Thales (c620–c546 BC). Aristotle (1998) called them *physikoi* ("physicists" or nature philosophers) to distinguish them from the *theologikoi* ("theologians"). Only fragments of their writings survive (Barnes, 1987).

and positivism. With new laws of motion on Earth and in space, the revival of atomism, an emphasis on new experimental techniques and instruments,[7] the creation of scientific societies and journals, including breakthroughs in mathematics (logarithms, analytic geometry, calculus)—these all helped establish physics as no longer a qualitative but an experimental and mathematical science. Yet, at the time, it was recognized as *natural philosophy* (Westfall, 1971).

But the new mechanical and reductionist worldview it was associated with—later referred to as the "clockwork universe"—carried deep philosophical implications for causation, materialism, and determinism, and was explicitly formulated to oppose Renaissance Naturalism and the scholasticism of the universities. Today it has become apparent that the so-called "pioneers of modern science" were not as "modern" nor as "scientific"[8] (Bowler and Morus, 2005, p. 24) as the popular conception of convergent realism (Laudan, 1981) would have us suppose; Or even as the *ahistorical*, positivist-influenced tradition in philosophy of science from the 1930s to 1970s—and still too many current textbooks—would have us believe. Interestingly, Bunge (1996, p. 317) held that "logical positivism remains the tacit philosophy of many scientists."

In our time, revolutionary shifts due to relativity and quantum theory have led to a critical reexamination of the *foundations* of physics and collapsed the preceding mechanical views (Einstein and Infeld, 1938; and Heisenberg, 1958). They reopened older philosophical doors looking into the nature of atoms, space and time, forces, and fields, while questioning the worth of classical causality,[9] determinism[10] and realism for troubling views regarding "uncertainty" and subjectivism (Bunge, 2012; and Romero, 2019).

Theories are not culturally isolated "pure" creations but are necessarily embedded in the philosophical background of ideas constituting their central elements. This background normally remains tacit but erupts when anomalies and alternative theories arise—forcing a reevaluation through *re-interpretation* and *disputes*—here philosophy explicitly comes to the fore. The renowned debate, for example, between Bohr and Einstein about the "Copenhagen interpretation" of quantum mechanics, was inherently deeply philosophical, concerning typically traditional questions about what can possibly be *known*, and what can be said about what possibly *exists* (Kosso, 1998): do scientific theories reveal actual truths about nature or are they just convenient fictions allowing calculations and predictions to be made that

---

[7] Telescope, microscope, thermometer, barometer, pendulum clock, air pump.

[8] One thinks of Kepler drawing horoscopes and Newton's decades long alchemical studies and the scrutiny of Biblical numerology. Dolnick (2011) adds that the Royal Society had many members that were charlatans—and included experiments using persons and animals that were ghastly. Most believed in witches and divine intervention. Unlike today, they avoided the stigma of "atheism."

[9] *Cause* is taken as prior to and responsible for the event while independent of the observer and mediated across space and time, whereas in special relativity the action is constrained within light speeds and the event "simultaneous" is observer dependent. The EPR paradox criticized quantum mechanics because it implies a violation of local (classical) causality, which the phenomenon of entanglement confirms.

[10] It asserts that the present state of a system and mechanical laws can determine with predictive accuracy the exact future state of the system (e.g., predict comet paths). But in quantum theory, there exists a non-eliminable indeterminism when calculating probabilities of outcomes due to inherent randomness in nature. Measurements are constrained by Heisenberg's uncertainty or "indeterminacy" principle.

"save the appearances"? In philosophy, this is called the quarrel between *realism* and *instrumentalism*; it runs straight through the history of physics to the present, but ironically, is unseen in textbooks. Neither this quarrel nor the Bohr-Einstein debate has been fully resolved—which may surprise students (Norris, 2000; and Becker, 2018).

Nor are such debates *new* or solely characteristic of 20th century science. Unlike typical textbook presentations whose exclusive focus is to show scientific consensus in retrospect for a triumphant theory, *interpretive* problems of theories, whether old or new, are never fully resolved (Sklar, 2000, p. 735). While consensus is the final goal of the physics community, *dissent* in fact is the life-blood of scientific growth: philosophers Lakatos and Feyerabend have argued that science stagnates without competing alternative theories.

The historical record is clear: controversies occurred throughout its progression, especially when established theories clashed with competing views: geo-centrism vs helio-centrism; Huygens vs Newton on the nature of light; Leibniz vs Newton on the nature of space, gravity and action-at-a-distance (Dainton, 2010); Ampère vs Biot on electromagnetism (Braga *et al.*, 2012); Fresnel vs Laplace, Biot, and Poisson on light as a wave (Kalman, 2010); Mach's instrumentalist view vs Planck's realist view of the nature of atoms, and not least, Bohr's willingness to abandon the laws of conservation of energy and momentum, and his principled stand against the photon idea (Pais, 1991, p. 233). One could go on.[11] "Scientific controversies are found throughout the history of science. This is so well known that it is trivial" (Machamer, 2000, p. 14). They represent in essence a sharpening of theoretical (philosophical) ideas and scrutiny of empirical evidence by exposing paradoxes, bias, and hidden or ill-thought out assumptions. They may occasionally even go hand in hand with "thought experiments," those instances where scientists appear especially imaginative and philosophical.[12] How can students learn about physics (or chemistry) *as practice* when teachers are uninformed about, and textbooks have mistreated, the actual historico-philosophical record? (Garritz, 2013). Here they could come to understand the nature of scientific *argumentation*, how experiments and data are adversely evaluated, critical thinking, and how disputes are rationally (though not always respectfully) resolved.

12.3 PHYSICS *AND* PHILOSOPHY OR PHYSICS *WITHOUT* PHILOSOPHY?

Given this record of the fruitful interaction of physics and philosophy, one is nonetheless surprised to come across physicists like Feynman, Weinberg, Hawking, Krauss and de Grasse Tyson, who either

---

[11] Ohm's law dismissed for the wrong Barlow's law, which held back electric telegraphy for decades; Thomson's dismissal of Rutherford's "atom;" Millikan versus Ehrenfest on sub-electrons; Einstein versus Millikan on the photo-electric effect; Duhem's dismissal of relativity; Chandrasekhar vs Eddington on gravitational collapse.

[12] Kepler's moon flight; Galileo's void; Newton's cannon ball satellite, rotating bucket; Maxwell's demon; Einstein's light beam traveler, accelerating space elevator and EPR paradox; Schrödinger's cat.

openly deride philosophy or dismiss it as useless (three of those mentioned obtained Nobel Prizes). It is remarkable to think that some well-known scientists imagine that philosophy can be gone around and wholly without.

Feynman's views attributed to him are familiar: "Philosophy of science is about as useful to scientists as ornithology is to birds" or "philosophers are always on the outside making stupid remarks." Hawking had opined that regarding traditional questions concerning the behavior of matter or the nature of reality philosophy has nothing to say, and in fact "… philosophy is dead. Philosophy has not kept up with the modern developments in science, particularly physics. Scientists have become the bearers of the torch of discovery in our quest for knowledge" (Hawking and Mlodinov, 2010, p. 5). Weinberg (1992) had argued in a chapter entitled "Against philosophy" that its influence in science can even be detrimental. Lawrence Krauss has stressed that philosophy is basically useless for the theorizing physicist. The former TV host of *Cosmos*, de Grasse Tyson, stated that the more we learn about the universe and quantum physics "… each of which falls so far out of what you can deduce from your armchair that the whole community of philosophers … was rendered essentially obsolete" (Rovelli, 2018, pp. 481–482). Yet taken together, such views have led one commentator to remark that the two fields exhibit "a love-hate relationship," given that there exist skeptical voices among some philosophers disparaging science (Koertge, 1998; and De Haro, 2020).

True, there is some albeit limited evidence that certain philosophical ideas have stultified scientific progress seen in historical hindsight, such as Descartes dualistic philosophy (Westfall, 1971) or when logical positivists dismissed the value of metaphysics, removing one key traditional aim of science, namely, the search for physical truths *behind* the appearances. In addition, there currently exists a debate among scholars as to what extent the tradition of *philosophical* atomism hampered more than it helped the rise of scientific atomism (Harré *et al.*, 2010). But surely these physicists are aware of philosophical ideas launched by scientists that hampered progress (e.g., epicycles, vortices, caloric, elastic fluids, phlogiston, ether, etc., Hesse, 1962)?

The fact that philosophers of science have disagreed among themselves about the nature of scientific method (e.g., Carnap, Popper, Kuhn, Lakatos, Feyerabend) has had little effect on scientific practice generally. On the other hand, the physicist Rovelli (2018) holds that vestiges of positivism, especially as later adjusted by Popper and Kuhn, have had some detrimental influence on theoretical physics, but not in the way the detractors of philosophy surmise. What the "anti-philosophy" scientists do not seems to grasp is that their own thinking is saturated with a particular, tacit philosophy of science (echoing Bunge):

> The current dominant methodological ideology in theoretical physics derives from their notions of *falsifiability* and *scientific revolution*, which are popular among theoretical physicists; they are often mentioned and are commonly used to orient research and evaluate scientific work … in declaring the uselessness of philosophy, Weinberg, Hawking … are in fact paying homage to the *philosophers* of science they have read, or whose ideas they have absorbed … The imprint is

> unmistakable. … [how] Tyson mocks philosophy, these criticisms are easily traced to the Vienna Circle's anti-metaphysical stance. … (p. 485, original italics).

These scientists seem unaware that they are caught in a self-referential paradox, since by denying the value of philosophy they are in fact doing philosophy—they are making an argument against its use. By using reasoning, they have become ensnared by philosophy herself. Moreover, the claim "science doesn't need philosophy" or "show me the evidence" assumes a rather neat separation between physics and philosophy, which has never been the case, neither in the past nor currently—as if scientific inquiry and reflection, and philosophical contemplation, do not intersect but occupy opposing realms of thought.

In reality, there has never been, nor can there be physics *without* philosophy. Science has always been conducted within the context of the philosophical ideas of the age. How can it be otherwise? Thinkers of any stripe are either children of the *Zeitgeist* or rebels against it. It is to say that scientists always think out from, and practice within given research programs (or *paradigms*, if you will) that are conditioned and constrained by their own history, language, and tradition—all themselves intricately imbued with philosophy in diverse ways, sometimes at their very foundations.[13]

The language shift, surely, is indicative of a transformation in the conceptual landscape, when pioneers like Kepler, Stevinus and Galileo argued that mathematics was the "language of science," acting like a hidden code as the abstract mirror of nature's constitution. This involved not just symbols, numbers and ratios but equally terms, principles and groupings of a new semantic *ontological* vocabulary. Thereto, common sense and direct observation (and Biblical texts) lost their status as sole legitimate criteria for discovering truths of nature, and science ever since has progressed on a counter-intuitive and observation-*defiant* basis. But clearly, the stress on mathematical representation and mechanistic philosophy revealed a new *ontological* outlook on nature, and this attitude persists to the present (both belong to the "metaphysical foundations of modern physical science,"Burtt, 1954). Following this line of reasoning, the *new* "language of physics" (formalism and interpretation) inaugurated with relativity and quantum mechanics in our time has likewise demanded another though distinctly different ontological outlook.

Other historical examples further illustrate the *interdependence* of physics and philosophy, e.g., the influence of the Platonic-Pythagorean tradition or the case of Giordano Bruno (Koyré, 1957; and Gatti, 1999). Not even the revolutionary ideas of relativity or quantum physics emerged from a philosophical-cultural vacuum (Torretti, 1999):

> Quantum mechanics springs from an intuition due to Heisenberg, grounded in the strongly positivist philosophical atmosphere in which he found himself: one gets knowledge by restricting oneself to the observable (Rovelli, 2018, p. 483).

---

[13] Sklar (1992, p. 9): "But in looking at what physics tells us about philosophical questions, we must always be careful to ask if philosophical presuppositions have been built into the theory itself."

Einstein himself was deeply influenced by the philosopher-scientists Mach and Poincaré, whose criticisms of Newtonian physics helped loosen the authoritarian hold the great physicist had on his mind, while groping toward his new theory.[14]

At minimum, and quite generally, philosophy provides

> …methods for producing new ideas, novel perspectives, and critical thinking. Philosophers have tools and skills that physics needs, but do not belong to the physicists training: conceptual analysis, attention to ambiguity, accuracy of expression, the ability to detect gaps in standard arguments, to devise radically new perspectives, to spot conceptual weak points, and to seek out alternative conceptual explanations (Rovelli, p. 484).

Given that philosophy can improve a scientist's analytical skill set, how else can the *relationship* between physics and philosophy be characterized? There is a common view that science progresses while philosophy does not, or that science has experimental methods for determining when its ideas or theories "go wrong," which aids in solving actual problems, whereas philosophy is preoccupied with a perpetual quest for "truth" but never attains it, nor solves its problems, despite centuries of endless (some might add fruitless) debates. Still, few doubt its value for developing a skeptical, critical mind-set.

Some have argued only where science *ends* then philosophy begins (Jeans, 1943); others emphasize that any apparent border is porous, even *continuous* (Zinkernagel, 2011). Some see differences in the nature of their subject matter and methodologies: philosophy's primary concern, as part of the Humanities, is on *verstehen* (understanding), whereas science is concerned with *erklären* (explaining) in terms of causal relations and reduction. "The difference in methodology is often summed up by the mantra: 'philosophy asks why-questions, science asks how-questions'" (De Haro, 2020, p. 309). But even such a contrast can be disputed (Ladyman, 2002). Yet, both disciplines happen to *converge* in current cosmology (Kragh, 2014).

While it is true that philosophy has focused on questions at the level of general abstraction (e.g., "what is the nature of being?"; "How is causation related to explanation and knowledge?"), the philosophy of physics as a *bridging field* between philosophy and science has focused instead on direct inquiries into natural phenomena linking such questions to specific theoretic concepts and problems. Furthermore, philosophy arises naturally when examining intrinsic theory puzzles: "interpretive difficulties … are internally generated by the theories themselves" (Sklar, 2000, p. 735).

One wonders how the above named five physicists would respond to the fact that many physicists themselves have acknowledged the worth of philosophy, and hence ask why so many have taken the time to write quite a few books on the subject? (Torretti, 1999). Some had even gone on to develop their

---

[14] Furthermore, he admitted the influence of such heavy-weight philosophers as Leibniz, Berkeley, Hume Spinoza, and Schopenhauer. Bohr was influenced by Kant, Kierkegaard, and Wittgenstein.

own philosophy of physics: Hertz (1857–1894) insisted on a *phenomenological* methodology, the view that "the job of physics was to provide a complete, accurate and simple *description* of the phenomena of nature, rather than an *understanding*, either in terms of models or in terms of principles a priori necessary or self-evident to the human mind" (Cushing, 1998, p. 367; italics added). Mach (1838–1916) influenced logical positivism, the view that "physical laws are simply convenient summaries of phenomena," with the rejection of all metaphysical concepts (e.g., "atoms") that transcend sense experience (ibid). Poincaré (1854–1912) and Duhem (1861–1916) expounded *conventionalism*. This doesnot imply that their philosophy was always sound, but it does indicate their need to philosophize.

More to the point, two prominent physicists of the 20th century (and one with a Nobel Prize) working within this science-based philosophical tradition have written books with "physics and philosophy" in their very titles! (Jeans, 1943; and Heisenberg, 1958). These two volumes are eminently accessible, and even today, any physicist or physics teacher will profit from their breadth and insight—albeit both books are biased by the Copenhagen standpoint and its congruent instrumentalist perspective. This view was later severely challenged by the *quantum philosophy* of Bell (2004).

Our arguments concerning the worth of philosophy parallel Heisenberg (1958, p. 161), who held that modern physics had forced to the surface long standing philosophical issues but now approached from fresh perspectives. Controversies have continued ever since. Indeed, "physics needs philosophy—philosophy needs physics" (Rovelli, 2018).

12.4 PHILOSOPHY OF PHYSICS (PoP)

The philosophy of physics as a field of inquiry into the *nature* of physics has already been alluded to in previous sections illustrating advances in the philosophical framework of physical theories since the first scientific revolution. What slowly became clear is that thinkers had struggled to understand what *kind* of philosophy the field of physics implied, or had accompanied, or was required for its further progress. Though a start was made with Bacon's (1561–1626) *empiricism*, Torretti (1999, p. 98) in fact takes Kant's (1724–1804) conception of Newtonian physics as "the first full-blown philosophy of physics," while Berkeley (1685–1753) is seen as the inventor of positivism. By the early 1800s, mechanical philosophy was in competition with romantic *Naturphilosophie;*[15] at the end of that century, with attempts to reduce mechanics to *energetics* or alternatively, to "electromagnetic mass" (Born, 1962; and Cushing, 1998). As mentioned, *phenomenalism* and *conventionalism* followed. As will become apparent, vestiges of *positivism* currently influence curricula (Becker, 2018). What is now evident is that there exists no all-inclusive single philosophy *of* physics. For education, such aspects and

---

[15] Though widely ignored today, Romantic science (ca.1770–1840) played a significant role in the "second scientific revolution" (Richards, 2002; and Hadzigeorgiou and Schulz, 2014). Kuhn himself (1977) lamented the lack of recognition of the novelty of that era (ideas, discoveries, institutions) among historians of science. Today, some recognition of that revolution is found (Bigaj and Wüthrich, 2015; and Watson, 2010).

viewpoints should no longer be tacitly held or overtly ignored; moreover, PoP's worth lies in enriching instruction and learning (Lange, 2002).

Within the last three to four decades however "philosophy of physics" as a *sub-discipline* within the philosophy of science has become more independent and created its own niche in the academy, often combining the research work of both philosophers and theoretical physicists. A handful of papers have been written which provide a useful overview (Zinkernagel, 2011; Kuhlmann and Pietsch, 2012; and Crull, 2013)[16]. This new sub-discipline has preceded those of others (e.g., philosophy of chemistry, of biology) that have equally staked out autonomous territory over roughly the same time period. But its "recent" inauguration should not misrepresent the fact that its concerns go back to the early 1800s (as discussed), as philosophers of physics freely admit.

This section will be preoccupied with an investigation of the sub-discipline by looking at some major topics and books defining the field. It serves as a road map for physicists or physics instructors wishing to pursue the various avenues for themselves. Suggestions will be made on which pedagogical levels (secondary, undergraduate, graduate) are best addressed and what mathematical background is assumed.

We start with the following: (i) *opening remarks,* (ii) defining *seven tasks* required of the sub-discipline, and (iii) examining how the field is *structured*, with a focus on key authors seminal to its development. The field ranges widely, as can be imagined, from historical-philosophical interpretive problems in classical and modern physics coupled to epistemological and metaphysical issues to research discourse at the "cutting-edge" of contemporary physics.

i. In general, PoP is concerned with the following:
    1. *interpretation* and critical examination of the assumptions, concepts and models of physical theories to ascertain what kind of understanding of nature is being projected, "… what they tell us about reality" (Lange, 2002, ix): E.g. gravitation understood as action-at-a-distance or as a field theory; whether or not quantum mechanics can be considered complete; whether or not string theory is vacuous as a framework for constructing fruitful, testable models.
    2. inspection of the *foundational issues* in fundamental physics, either in, e.g., classical dynamics or electromagnetism (Batterman, 2013; and Sklar, 2013) or more commonly, problems currently preoccupying physicists and their modern "core theories": relativity, quantum and statistical mechanics (e.g., Knox and Wilson, 2021).

    Though the roles physicists and philosophers take on can play off in various ways, they can consist of co-operation in areas of advanced research (e.g., Nobel laureate Gerard' t Hooft on the conceptual basis of quantum field theory or Rovelli on quantum gravity; Callender and Huggett, 2001; and

---

[16] Muller (2023) lists the key authors, now in the third generation, numbering between 100 and 250, not including Ph.D. students and physicists. As of July 2020, the Pittsburgh Archive has listed 2882 philosophy of physics items, dominated by Quantum Theory (1673) and Relativity (756).

Butterfield and Earman, 2007), or by friendly critique and rivalry: "Philosophers of physics don't just accept what physicists tell them; a large part of their job is to interpret the constructions of physicists (and thus go beyond the ability of such constructions to yield accurate predictions)" (Rickles, 2008, p. 5). Then, again, this belief may not fully define the distinction between their two roles:

> After all, physicists do not merely derive predictions from their theories: for example the primary aim of string theory is unification and not empirical predictions. Indeed, much of physics seems to be about making abstract claims regarding the structure of the physical world. However, there is a difference in that physicists are often more pragmatic than philosophers of physics in accepting physical theories when they are successful and progressive (Kuhlmann and Pietsch, 2012, p. 210).

ii. *Tasks*: The two authors had originally listed *eleven theses* that they believe define the *tasks* of the sub-discipline. Here we only front *seven* tasks: exploration of three main issues involving methodology, fundamental concepts and ontology; contributing to physical knowledge (admitting in this case the line between physics and philosophy is "blurry"); engaging with inquiries similar to when physicists are in periods of crisis; explicitly going beyond the purely mathematical framework of theories; incorporating a more pluralistic and non-partisan approach to theories and concepts;[17] remaining interested in non-fundamental and abandoned theories; and linked thereto, staying historically informed "… the historical ignorance and the frequent reference to historical pseudo-accounts in physics textbooks is notoriously lamented" (ibid, p. 212).

iii. *Structure*: Of *what* does the philosophy of physics comprise? These *tasks* can be explored in narrower or broader contexts, as expected, depending upon how the subject matter is divided and which of several viewpoints are taken:

   - *Core theories:* one popular account would see the sub-discipline divided into *three* distinctive areas, seen as *three major pillars* consisting of the "core theories" of modern physics: relativity, statistical mechanics and quantum mechanics (QM).[18] Together these are taken to underlie all other phenomena of nature. Sklar's (1992) *Philosophy of physics*—the *first* to explicate the field—had organized his by-now "classic" (and non-mathematical) introductory book according to these theories and their themes. A recent textbook by Rickles (2016, p. 18), also written for the undergraduate, likewise draws attention to the same three "pillars" with the caveat that they can be misleading. Yet, the division persists in current handbooks reporting on recent research (Muller, 2023). These are more appropriate for graduate students (e.g., Butterfield and Earman, 2007; and Knox and Wilson, 2021).

---

[17] They write that physics tends to be dogmatic and orthodox about its basic assumptions and neglect alternatives, which PoP welcomes, like Bohmian or Everettian quantum mechanics. Becker (2018) has provided damning criticisms regarding the treatment of researchers in the past (e.g., Bohm, Bell, Everett, Clauser, Zeh) who questioned Copenhagen orthodoxy and insisted on working on problems in quantum foundations, including threatened career stagnation (before it became slightly more respectable by the 1980s–1990s; DiVincenzo and Fuchs, 2019).

[18] For popular accounts, see "Philosophy of physics" in *wikipedia* or Albert, *Encyclopedia Britannia*.

- However, Lange (2002, p. xvi) warns such a tri-partite grouping emphasizing the latest research findings reinforces the false impression that philosophy only has something positive to contribute after the "real work" of the scientists has ended—echoing Jeans: "The most harmful consequence of this approach is that it makes the questions investigated by philosophy of physics appear "merely philosophical" in a pejorative sense: marginal, detached from the concerns that actually drive innovation in physics."
- *Historical approach*: another account takes a broader historical perspective, e.g.,Cushing (1998) and Torretti (1999). The former has focused on the historical relation between philosophy and physical theories starting in Greece, up to debates of the Newtonian worldview and relativity, and ending with disputes in quantum mechanics (e.g., Copenhagen vs Bohm's pilot-wave theory, Bell's theorem, etc.). Missing is the statistical mechanics "pillar." A similar strategy is found in Torretti, but starting with Galileo. While Cushing can be a resource for students, Torretti is a magisterial study (including calculus). Both books delve into philosophy of science topics (inductivism, empiricism, approximate truth, incommensurability, underdetermination).
- *Fundamental concepts*: a third viewpoint is preoccupied with essential concepts, as exemplified by Lange's textbook (2002). It too, is broad in scope, but concentrates instead on basic terms and topics familiar to most students from their compulsory textbooks, typically, investigating mass, force, energy, electricity and magnetism, etc., but scrutinized with a philosophical lens as to their essence (using mathematical but non-calculus approach). Thick with arguments, laid out with problems, discussion and thinking questions, it explicitly encourages students to *philosophize for themselves* on topics usually presented to them solely as numerical problems but not philosophical dilemmas.
- *Narrow accounts*: focused either on *one* pillar (e.g., QM: Bunge, 1973; d'Espagnat, 2006[19]; and Maudlin, 2019) or *two* (e.g., relativity and QM: Kosso, 1998). Kosso is recommended *for students* for his remarkably clear exposition of the two theories and their link to epistemological vs metaphysical realism/antirealism. This dispute is rarely addressed in courses: whether physical theories deal with truths of nature (i.e., their postulated entities, such as quarks, are *real*) or are merely "empirically adequate."[20]

Bearing in mind the anti-philosophy comments of our physicists (above), and reflecting on the aforesaid books and handbooks, one can rightly conclude that philosophy is neither "dead," nor that philosophers have not kept up with the latest developments, nor that it is "useless." Quite the opposite, it seems physicists have not kept pace with the expanding philosophical literature contributing to contemporary

---

[19] The late CERN physicist Bernard d'Espagnat's weighty tome fills 500 pages. The physicist-philosopher was a student of de Broglie, worked under Fermi and Bohr, but belonged to a long line of those who were harsh critics of Copenhagen school, including de Broglie, Einstein, Schrödinger, Bohm, Everett, Bell, Bunge, Cushing, Penrose, Shimony and Smolin.

[20] This term means that a theory's *observational* predictions are correct; it leaves unanswered whether or not postulated conceptual entities exist (e.g., energy bonds, potentials, wave function, etc.). All true theories are empirically adequate, but not all empirically adequate theories are true. Many theories have made correct predictions and explained data, but were later shown to be false (e.g., Ptolemaic astronomy, phlogiston chemistry, Newtonian gravitational theory, Maxwellian electrodynamics, and old quantum theory).

conversations for decades. "How have physicists failed to get the memo from philosophers after all this time? Part of the problem is that physicists generally donot know much about philosophy. There is a massive asymmetry between the two fields …" (Becker, 2018, p. 272). A notable exception was Margenau (1978).

12.5 PHYSICS, PHILOSOPHY, AND THE FIELD OF EDUCATION

Regardless of all these relevant aspects, useful books, and evidence of philosophy *in* physics, it must be confessed that the contemporary relationship between the two disciplines is not at all evident to many in the educational field. This fact becomes noticeable in any casual conversation with either physicists or physics teachers. Not uncommon are questions like "what does philosophy have to do with science?" or more succinctly and less pejoratively, "how can any sort of "philosophy" contribute to helping my students better understand difficult physics' problems?" The interrogative presumes an obvious lack of evidence. Such questions implicitly assume of course a deep divide between science and philosophy, and certainly between physics education and philosophy. While instructors need not be openly hostile to philosophy, they certainly appear indifferent. Will things change with the recent arrival of PoP courses and textbooks (e.g., Rickles, 2016) or new quantum foundation textbooks (e.g., Norsen, 2017)?

12.5.1 Philosophy of physics: Reforming textbook pedagogy and instruction

Nonetheless, the outlook that physics has no need of philosophy (as held by many today) does have a number of harmful consequences for *physics education*. Two immediately come to mind and combined they discourage philosophical reflection of learners: the weight placed in textbooks on *operational definitions* of key terms (over *conceptual*),[21] and the attitude toward students' curiosity, imagination and questions during instruction. These two actually bear witness to a hidden positivist-type PoP among teachers and curricula.

Examining the first one, operational definitions are vestiges of logical positivism and demarcate a quantity by telling us simply how to measure it—focused on instruments with defined units—e.g., "time is what a clock measures." One forgoes the need to delve into its *meaning* (Bunge, 1973, p. 10: "…far from assigning meanings, measurements presuppose them"). One can fairly claim that curricula exhibit *operationalist bias*. Any examination of older or newer textbooks makes this quite evident.[22]

---

[21] From the Nobel Laureate Bridgman in the 1930s, who was influenced by Mach's positivist argument.

[22] Textbooks of the 1970s and 80s were surprisingly open about the positivist legacy (see, Marquit, 1978). The penchant today is either to ignore the heritage when the term is employed (Giancoli, 1995; and Young and Adams, 2020) or to overlook the term entirely while utilizing the idea (Halliday *et al.*, 2013; Knight, 2017; and Cutnell *et al.*, 2018).

(Holton and Brush's textbook addressed the pros and cons of this bias). While operational definitions may occasionally be necessary, they are hardly sufficient.

Why is this problem? Because any direct linkage to foundational (metaphysical) issues disappears by robbing *all central concepts* of their underlying depth. It presents another obstacle blocking students from achieving an enhanced conceptual insight (Kalman and Lattery, 2019). Avoidance is universal: textbooks regularly evade asking *foundational questions* when first introducing them, i.e., "what is space?" or "what is time?"—instead, the text quickly links the concept to a measuring instrument. *Time* is utilized in most equations as an afterthought ("time interval"), and any conceptual notion is lacking. Likewise, "space," as to its *nature*, makes no explicit appearance, but these same books show no embarrassment in a relativity unit where short talk about "spacetime" suddenly appears. At best "space" is allied with distance and length measurements.

If the original Newtonian *metaphysical* framework is not even presented (with concepts of absolute space and time, and the shift from Euclidean space to Minkowski spacetime) how is the student supposed to recognize the *discontinuous* shift or tremendous conceptual leap? (Levrini, 2014). They cannot of course—worse, such an exposition contributes to physics folklore that knowledge is built in a continuous, non-controversial and convergent way to "truth". A similar slight of hand can occur in the exposition on wave-particle duality (in fact a *paradox*, where any sense of "crisis" is missing), seen rather as a curious "oddity" of nature (Greca and Freire, 2014).

In all three cases, the philosophical "edge" has been removed, the mental "knife" has become dull:

> High school teachers and textbooks transmit an incorrect image of science, which ignores the existence of crises and paradigm shifts. The introduction of topics in modern physics … [occurs] without reference to its essential novelty or to the main differences between the classical and the new paradigm. A suitable occasion for showing the richness of the development of science and importance of … revolutions is thus wasted (Gil and Solbes, 1993, p. 260).

Freshman textbooks hardly fare better. Kuhn (1970, p. 136) blamed textbook writers who abuse history and make revolutions "invisible." Little has changed in 50 years.

One can go on: "force" has been notoriously hard to define (Sklar, 2013): most textbooks today describe force as "the cause of acceleration" (the second law), yet Poincaré stressed "to say force is the cause of acceleration is to do metaphysics" (Coelho, 2010, p. 102). Other central concepts have not fared better: "mass" or "field" is usually related to equations where calculations can proceed, measurements made or to general laws: *mass* is usually defined in terms of the law of inertia based on Mach's positivist influence and operationalism (Hecht, 2006, p. 41), though this popular definition is often *circular and competes with others* (Roche, 2005); "Field" is often not in the index (nor "space")—it appears in sections restricted to calculations. Is it a real entity, or merely a convenient calculation tool, a fiction? This results when removing the *ontology* behind physics' concepts (Einstein, 1934). Essentially, textbooks stress viewing physics "as the search for phenomenological correlations, rather than the essential properties of matter which manifest themselves in the phenomena" (Marquit, 1978, p. 787).

The upshot means philosophy has no opportunity to enter the class discussion at the introductory level of basic questions concerning physical ideas, never mind its *laws* and *theories*. For the latter two, even such *nature of science* terms may not be explicated (in an introductory chapter), along with others like "inference," "hypothesis," "modeling," "confirmation," and even "idealization"—all at the core of what constitutes scientific explanation (Ladyman, 2002).

If learners are never confronted with fundamental ontological-type questions ("what is space?;" "what is energy?") or other epistemological-type questions ("how do we know "X"?; "has "X" been confirmed?"), related to the many essential ideas that fill their textbook pages, how are they expected to develop *critical thinking* (see Arons, 1997, "Critical thinking")? (This can be done if the instructor opens a new topic with a question, rather than an equation entailing an answer, and employs an appropriate strategy, like *think-pair-share* or *Socratic method*). If never presented with the opportunity to mull over concepts first, or better, to ask *foundational questions*, how can one expect them to develop better understanding? Or expand their imagination? (see Sec. 12.6). Worse, there is little room for their curiosity to be satisfied when the awe and mystery of physics is lost.

Research has shown that problem-solving, improving manipulation skills and memorization alone are substantially inadequate (Kalman, 2017; and Schulz, 2019). How will an instructor respond when asked: "are fields real things?" or "what exactly is charge?" and "how can a mass particle like an electron have a (de Broglie) wavelength?" (As an answer, is Bohr's "complementarity principle" adequate? Does it resolve the issue?) Will such student queries be dismissed as "philosophical," hence can't be answered and irrelevant? Or they distract from "getting on" with the course material?

How will students respond when they discover "there is no really good definition of mass"? (Adler, 1987; and Hecht, 2006). Here is a concept at the very heart of Newtonian *and* relativistic physics (Einstein, 1954; Lange, 2002, pp. 224–250; and Roche, 2005). In addition, students are often quite astonished after spending hours calculating energy-type problems in their textbook when given the Feynman quote: "It is important to realize that in physics today, we have no knowledge of what energy is." (Thereto, is potential energy *real*? Hecht, 2016). Philosophy is not far below the surface in the physics classroom, and can only be addressed if the class is "slowed down" to allow for discussion, and textbooks are revised accordingly.

Such elemental questions regarding subject topics can be multiplied, obviously, for the dozens of concepts introduced. Yet, their textbooks for the most part will not help them, as un-philosophical as they are. Those students who take physics courses because they see themselves as seekers of deep truths of nature will be disappointed and discouraged (Smolin, 2006; and Hadzigeorgiou and Schulz, 2017).

We come to the *second harmful consequence* regarding students' curiosity and concerns in the learning atmosphere. We note that while physicists were dismissing philosophy in general, some of those involved in physics education were drawing contrary conclusions, even as early as the 1940s (Frank, 1947; and Eger, 1972). They drew attention not only to the significance of philosophy for improved learning but also to the fact that some philosophies can be inherent in not *hidden* in textbooks and teacher pedagogy. Sadly, the admission by Seeger (1960, p. 385) that "every physics teacher, in short, teaches some form of the philosophy of physics, either directly or indirectly" has remained unrecognized. This state of affairs can surface in ways not admirable:

> … it was almost as if we were being taught to look down on people who thought about foundational problems. When we asked about the foundational issues in quantum theory, we were told that no one fully understood them but that concern with them was no longer a part of science. The job was to take quantum mechanics as given and apply it to new problems. The spirit was pragmatic: "Shut up and calculate" was the mantra. People who couldn't let go of their misgivings over the meaning of quantum theory were regarded as losers who couldn't do the work (Smolin, 2006, p. 312).

Lange (2002, p. x) also bears witness to this stance. Whenever the topic of quantum mechanics is broached, the educational failure (avoiding its intrinsic metaphysical problems) becomes absurdly apparent (Greca and Freire, 2014). Textbooks at both secondary and tertiary levels almost universally, and quite astonishingly, proudly proceed as if no such quandaries exist, instead keeping to a hardnosed problem-solving account.[23]

One imagines the eagerness with which classical physics was taught in the 19th century, likewise ignoring foundational problems until Hertz, Mach, Duhem and Poincaré fully exposed them (Sklar, 2013). Unlike the prescribed textbooks today, which has many people convinced by their composition and exposition, genuine *interpretative* problems in the foundations of classical physics abound, but remain deftly buried—the dilemmas are by no means restricted to modern physics (Lange, 2002; Coelho, 2012; and Hecht, 2016). Such a narrow educational emphasis does not contribute to critical reasoning, nor does it excite students to a sense of wonder or curiosity.

Questions concerning the metaphysics of physics cannot really and should not be avoided, though physics education culture repeatedly attempts it. If metaphysical inquiries are purportedly so irrelevant to both physics and therefore physics education, why then, one can ask, are they so vital today at the research edge concerning the nature of causation and physical laws, notions of time and quantum non-locality, teleportation, quantum cryptography, cosmology, and others? (Maudlin, 2007; Lam and Esfeld, 2012; and Bigaj and Wüthrich, 2015). Some have even argued that the relationship is symbiotic (Belot and Earman, 1999). "A society which is uninterested in metaphysics will have no theoretical science" (Hesse, 1962, p. 303).

12.5.2 Philosophy of physics: Learner motivation and interests

Unlike the research domains of many practicing physicists, the field of education brings with it not just its own epistemological issues but also inexorably *ethical* ones. Instructors certainly must be more

---

[23] There is no mention whatsoever of the "measurement problem" when discussing Ψ (e.g., "collapse"), although it has been unsolved for 90 years. Also, unseen are Bohmian pilot-waves, spontaneous collapse theories, Bell's theorem, or Schrödinger's Cat (first mention of quantum entanglement). Norsen (2017) addresses all of these. See Cordero (2019) for philosophers' analysis.

cognizant of what philosophy of physics *they* employ or even possess unconsciously. That means they require an examination of their own *personal* philosophy of physics education (Schulz, 2014a, 2014b): they will need to make value judgments on how important they think historico-philosophical issues are and what impact they could have for deeper learning (Cushing and Lange have shone significant light here). Are they willing to "make room" for the history and philosophy of physics in classrooms to engage students on that basis? In other words, are they willing to engage with the *ontology of physics*, past and present? What should the *ultimate aim* of introductory physics courses be?

Philosophy performs another critical function concerning the *meta-evaluation* of the nature of physics education itself: one should ask the key *curriculum question* whether the makeup of courses and textbooks at secondary and introductory levels are appropriate for the intended audience, being deliberately designed as *technical pre-professional training* (TPT)? Here we come back full circle to the themes in the *Introduction*.[24] TPT is not just a pedagogical attitude; it serves as a (limited) *educational philosophy* (Schulz, 2014a). Would it not be wiser to provide students with not just technical expertise—where too many "cannot see the forest for the trees" once the course is over—but equally with a broader conception of physics? (see Matthews 2015, p. 62).

One can examine learners' motivation and engagement on a spectrum. At two extremes, simplified, there are the "seekers after truth" and the "pragmatic" seekers wishing only to fulfill course requirements to pursue the next level of their Bachelor program of choice. For the first group, the question of ontology and truth is paramount and the primary motivator, and problem-solving is of secondary importance; for the second it is reversed: satisfaction consists of finding correct answers to puzzle-like math manipulations, the ontology of secondary worth, if not irrelevant. While TPT culture will inevitably do injustice to the first group, it somewhat *misleads* the second about the actual nature and development of physics, as illustrated.

This disparity can be explained heuristically by the fact that the two groups are involved with two different kinds of *meaning making*, and opposite thinking styles: the "narrative" mode vs the "propositional" (logico-scientific) mode, according to the psychologist Jerome Bruner. The first construes particulars in normative contexts like stories (e.g., scientist's struggle for knowledge; controversies), which provide coherence and generality to meanings, where interpretations can vary; the second is dominated by rules of logic in formal symbolic and conceptual systems, held to be autonomous and context-free, allowing only unique solutions (correct or incorrect) and no alternatives. "Logico-scientific" thinking is primarily impersonal and convergent, while the first is personal and divergent (see, Hadzigeorgiou and Schulz, 2019). TPT culture in its preoccupation with the second has no actual pedagogical sight of the first (Höttecke and Silva, 2011).

---

[24] Currently worldwide, TPT strictly serves as an induction into academic physics at post-secondary—although it is widely accepted that the vast majority of students taking such courses (over 90%) will never enter a Bachelor's program in physics, but instead a professional program in engineering, chemistry, education, etc.

12.5.3 Philosophy of physics: Aspects for improving instruction and curricula

In this subsection and the next, we spotlight key PoP questions and topics that should be integrated into introductory physics classes to enrich subject material, enhance conceptual understanding and develop critical thinking. Instructors willing to shift the pedagogical culture are not expected to engage in lengthy philosophy of science topics in classrooms, but rather include these items when subject topics are broached in curricula to the degree required as to age, ability and level (secondary or post-secondary). It also represents a challenge to authors of introductory textbooks to revise relevant sections to encompass or "make room" for the aspects listed, to emphasize the value, and expose students to foundational questions and problems. (While most remain unsolved or in dispute, their open nature encourages curiosity and wonder.)

There clearly exists a surplus of books on the philosophy of space and time, the philosophy of relativity or quantum mechanics, etc., some quite advanced, which have purposely not been identified here. The specific PoP resources cited were thought sufficient to aid in this endeavor, especially those interested for the first time.

- General
    - What is the nature of physical laws? Are they mere regularities? Are they approximate, or idealized, or do they lie? (Cartwright, 1983)
    - What is the epistemological role and linkage of models, laws and theories?
    - Can derelict theories (like Newton or old quantum theory) still possess "approximate truth"? Is the aim of physics "truth" or mere empirical adequacy?
    - Do "crucial experiments" truly exist in physics, and if so, what are the criteria?
    - What is the relation between conservation laws and symmetry? (Feynman, 1997; and Rickles, 2016)
    - Why do the fundamental constants have their values? (m_p, m_e, e, G, h, α)
- Space, Time, and Relativity
    - Are space and time entities that exist apart from atoms, planets and galaxies? (Sklar, 1992; and Maudlin, 2012).
    - What is the nature of time? (Theories of time: illusion vs presentism vs arrow; here, discussion can overlap with the famous "twin paradox" of SR and provide context; see Dainton, 2010).
    - Is space best seen as absolute or relational? [The classic debate between Newton and Leibniz (Smart, 1964) deserves a place in curricula considering the modern version of substantival vs relational views are relevant to current debate in GR; see Dainton (2010); Levrini (2002); and Hoefer (1998).]
    - Are Newton's and Einstein's theories of gravity incommensurate?
    - Can we know the true geometry of the universe? [Discussion of Euclidean vs non-Euclidean geometries where curricular topics move from SR to GR (Kosso, 1998; and Dainton, 2010).]

- Thermodynamics
    - Puzzle of time's arrow: does the entropic increase (*asymmetry*) represent an asymmetry in time itself? Do we infer/reduce the former from the latter or vice versa? (Price, 2006)
    - How can many observed macro-phenomena with entropic asymmetry be established from laws of dynamics at the micro-level that show no such asymmetry [i.e., allowing time reversal? (Sklar, 1992, 2000; and Besson, 2014)]?
- Quantum Mechanics
    - Does the "underdetermination of data" thesis adequately account for the great empirical success of the old quantum theory? What does this imply for realism? (Pais, 1991, p. 269)
    - What is the role of classical concepts in the quantum mechanical description of the world? (Sklar, 1992; and Kosso, 1998)
    - What is the nature of the quantum state described by the wave function? Does it reflect some aspect of the real world or is it only an intermediate calculating device? (Cushing, 1998; Maudlin, 2019; and Cordero, 2019)
    - In measurement, does the "collapse of the wave function" reflect a real *physical* phenomenon? Is it a coherent interpretation? (Cushing, 1998; and d'Espagnat, 2006)
    - If the Copenhagen interpretation is not monolithic, are its versions coherent?
    - Is the Bohmian "pilot-wave" theory a viable counter-interpretation to Copenhagen? What other interpretations might be viable? (Rickles, 2016; and Maudlin, 2019)

12.6 PHILOSOPHY OF PHYSICS AND PER: LEARNING THEORIES AND INSTRUCTIONAL STRATEGIES

Many students do not conceive of scientific knowledge as a highly ordered, coherent, knowledge structure that contains a set of interrelated ideas. This is part of the barrier that prevents them from accepting and understanding important physics concepts and laws, e.g., classical mechanics. They also harbor misconceptions about the *nature of science*. Having them examine the philosophy of physics can help them overcome these barriers, develop critical thinking and better understand how science progressively develops.

12.6.1 Student Epistemology: P-prims, "Theory-Theory," nature of science

Physics education research has described students' knowledge structures from two different perspectives: the so-called "Knowledge in Pieces" (KiP) view and the "Alternative frameworks" or "Theory–Theory" (TT) view. The first, as described by di Sessa (1993 and 2008), holds that a student's personal knowledge is fragmented or "in pieces": it consists of isolated structures called *phenomenological primitives* ("p-prims") which are activated depending on the situation studied or analyzed—thus, they are situation-dependent. The second, based on the influential early paper by

Posner *et al.* (1982), maintains that students enter classrooms with largely stable and coherent ideas about the natural world but that can differ substantially from those presented in science textbooks and lectures—thus functioning as naïve theories. |Vosniadou *et al.* (2008) also present an alternate view of students' knowledge of science called framework theory. This is seen as consisting of basic "presuppositions" about how physical bodies function in the world. On this basis knowledge acquisition is a gradual process during which existing knowledge structures are slowly revised, not as radical as Posner *et al.* proposed. The seminal work of Chi (2013) places physical entities in *ontological* categories based on which they offer a definition of conceptual change. She explicitly claims that naïve theories and scientific theories are often *incommensurate*. Posner, Vosniadou, and Chi were all encouraged by the philosophy of science.

The idea of "incommensurability" stems from two philosophers of science, specifically Kuhn (1970) and Feyerabend (1993). Kuhn used "incommensurable" to mean the *holistic nature* of the changes taking place in a scientific revolution. For example, Newton's theory was initially widely criticized because it didn't explain the cause of attractive forces between matter, something required of any theory of mechanics from the rival Aristotelian and Cartesian theories. Such developments require replacing existing concepts with new concepts that are incompatible with older ideas. Feyerabend (1993, p. 212) stated: "In 1962, I called theories such as those containing [medieval] "impetus" and "momentum" incommensurable theories." Here his incommensurability corresponds to questions and concepts that have meaning only in a particular theory framework. However, if an *overlap* exists between successive theories (i.e., shared ideas and/or concepts), meaningful questions *can* be asked in the context of both theories. For example, within both the wave and particle theories, we can ask the question: Does light bend around an obstacle (diffraction of light)? But, where there is no overlap, there exist questions that are only meaningful within the context of one theory but have no meaning whatsoever in the context of another. Thus, a question on the nature of the *ether* is meaningful in the context of the wave/ether theory but is meaningless in special relativity.

Identifying and addressing student ideas that are incommensurate in important respects with canonical (target) scientific knowledge remains a significant challenge for science education. This harkens back to the Posner *et al.* (1982) thesis that conceptual change by students is *analogous* to theory change in science as described by the two philosophers. Chi wrote

> Although students can readily learn by adding new beliefs about "internal force," such as the equation for its relation to mass and acceleration, the definition of acceleration, and so on, these newly added beliefs cannot correct a student's conflicting belief that a thrown object acquires or contains some internal force [i.e., impetus theory]. Moreover, such conflicting beliefs cannot be easily denied or refuted by contradiction. For example, stating that "a thrown object does not acquire or contain internal forces" or stating that "a thrown object contains some other kind of forces" will not succeed in helping students achieve correct understanding. … We propose the operational definition that certain misconceptions are robust and difficult to change because they have been mistakenly assigned to an inappropriate "lateral" category (p. 51).

Entities are objects or substances that have various attributes and behave in various ways—a ball is a physical object with attributes such as mass, volume, shape, and behaviors such as bouncing and rolling. But many students view force as a *substance* kind of entity that can be possessed, transferred, and dissipated. Students often explain that a moving object slows down because it has "used up all its force" (McCloskey, 1983), as if force was like fuel that is consumed. We claim this is why some misconceptions are so robust – because the naïve conceptions are *mis-categorized* into an ontologically distinct tree. When students' misconceived ideas conflict with correct ideas at the lateral category level, then refutation at the belief level will not promote conceptual change. This is because refutation at the belief level can only cause local revisions of the features/attributes/values of certain dimensions, whereas conceptual change of category mistakes requires changing the dimensions, which may require a categorical shift.

Another chief misconception of students regards their view of the *nature of science* (NOS) and how it develops. There is a major distinction to be drawn between a valid scientific theory and the discovery of regularities in nature. Regularities that occur through discovery learning are basically Baconian in character. Francis Bacon argued that all scientific knowledge was based upon careful observation and inductive reasoning (Bacon, 1863, Book 1, CXVII). Hence, the starting point of all sciences is experimentation, and regularities or patterns found in events by experiment or observation would reveal *laws* (causes and axioms) and these would lead to further experimentation. The natural regular underpinnings of the world (laws) are revealed only in this way. Implicit in this view is that any "theory" in science can be justified only if it has been deduced from the performance of experiments. Moreover, many textbooks tacitly tend to present science in such an inductive fashion, and many students believe that science proceeds primarily in this way, which is incorrect (Bauer, 1992).

Such "discovery learning" has its merits, for it is ideally suited to search for regularities in nature, as examples are the discovery of Titius-Bode law, Mendeleev's discovery of the Periodic table of the elements, and Gell-Mann and Ne'eman's discovery of the eightfold way in 1961 (that led to the quark model). In each case, the law revealed missing "elements" that were found by later observations. But occasionally a regularity like the Titius-Bode law is fictitious—because in that *specific* case it gave no hint toward theory. A law of itself may or may not advance theoretical physics; while most laws show regularities, not all regularities are laws. (Laws are further restricted to the narrow domain where they apply). Balmer (1825–1898), as an example, discovered an *empirical formula* that matched four lines in the visible hydrogen spectrum. Afterwards, many other lines predicted by the formula (regularity) were found in infrared and ultraviolet spectra—but an *explanation* of the success of the formula had to await Bohr's atomic theory. Yet, this basically encapsulates Bacon's scientific method: perform many experiments, examine them for regularities, and formulate laws, leading to further experimental predictions to be verified. Certainly, the idea that a regularity in nature yields a law is Baconian, and one basis of *empiricism* as a PoP (Cushing, 1998).

That this perspective on science rapidly became the norm is clear from Newton's famous statement in his work setting forth the laws of mechanics: "Hypotheses *non fingo*": I suggest no hypotheses—when asked what caused gravity. Newton, despite his prestige, was concerned about accusations of making theoretical pronouncements that did not fit with Baconian science. In actual fact, Newton made hypotheses throughout his *Principia*. It was later recognized, in contrast to the inductivist-empiricist oriented philosophy of physics, that another philosophy proceeding from theory and hypotheses through to experimental disconfirmation, or *hypothetico-deductivism* and *falsification*, were equally of use, and better, according to Popper (1963). Moreover, a high-level theory cannot be derived solely from induction because data usually remain theory-laden and underdetermined.

12.6.2 Instructional Strategies for developing a critical mind-set

12.6.2.1 Using philosophers of science and collaborative groups

It is instructive for students to compare and contrast different philosophies of physics. Specifically, studying the five philosophies of Bacon, Popper, Kuhn, Lakatos and Feyerabend not only helps students develop a more coherent view of science but also further develops their critical thinking skills. Philosophers' perspectives on science exposes how scientists have come to examine their views, which in turn can help students come to examine their own ideas on science, such as how theories change and knowledge is acquired. By the same token, when students examine various philosophies of physics, they can become aware of different ways of viewing the same content material. How this can be successfully achieved in an actual classroom will be illustrated by concrete research examples (Kalman, 2002, 2010, 2017).

Using collaborative group work, three to five students were organized with the task of viewing physics content through the eyes of a philosopher throughout the course. In doing so, they must come to understand and critically analyze their own personal views. Only then can they examine the evolution of science and develop correct ideas about how science advances. The students present these ideas to the class and additionally hand in a written version. (Only the written version is marked). The development of the written version is crucial in the development of the student's critical thinking skills. Marzano (1993, p. 155) notes

> By definition composing is a highly complex task that includes such phases as planning, translating and reviewing, all of which require a great deal of conscious control. … The longer the process continues and the longer the transcript becomes, the greater the interdependency among decisions. In short, the process becomes one of making decisions based on increasingly more numerous and complex conditions.

Students are challenged to produce increasingly more complex expositions based upon input during the oral presentations and written feedback on the written versions.

- for the first presentation, the instructions to the students were "Introduce your group's philosopher and explain the epistemology and methodology of your group's philosopher." For the remaining

group presentations, students are asked to explain how particular scientific developments would be viewed by their philosophers.
- the instructor throughout these presentations acts to suggest further topics for exploration in the next presentations. For example, what is Popper's attitude to *verisimilitude*?[25]

In an introductory calculus-based course on optics and modern physics, the changing historical attitudes toward the nature of light were explored. Various developments connected with light are presented to the class, explained in terms of light as a particle, light as a wave, and light as consisting of photons.

From Newton's time through to the end of the eighteenth century, it was generally accepted that light was made of particles. This was based not only on the authority of Newton, who stated this view as opposed to a primitive notion of wave theory proposed by contemporaries Hooke and Huygens but also on a careful experiment (1727) on stellar aberration by James Bradley (1692–1762). It could easily be explained by the particle theory of light, but not wave theory, until the concept of energy was fully developed later in the nineteenth century. By then, however, other experiments, conversely, could not be explained by particle theory but instead by wave theory—as developed by Young in 1801 and Fresnel around 1815. Here are examples of experimental *underdetermination*. In 1905, Einstein showed that the "photoelectric effect" (an anomaly first observed in Hertz's 1887 experiment) could be explained by light acting as particles impacting on a metal (originally highly contentious).[26] Subsequent to this paper, the view was adopted that light was neither a particle nor a wave, but paradoxically showed aspects of both but never at the same time, dependent upon the corresponding experimental design. This phenomenon was eventually called "wave-particle duality."

The exact topics explained by the students were

1. Discuss the wave and particle theories of light from the point of view of your philosopher. (You might not find these views in a book—it is a problem for you to solve!) In particular, comment on the role of Young's experiment as a *crucial experiment*.[27]
2. Discuss the ether theory (pre-1880) from the point of view of your philosopher.
3. Discuss the Michelson-Morley experiment from the point of view of your philosopher.
4. Discuss the existence of photons from the point of view of your philosopher.

Essentially, the class presentations were given in a manner to stress the components that make up a "good" theory and that scientific theories can change. But what constitutes a "good" scientific theory?

---

[25] Taken to mean "approximate truth or nearness to truth" (Popper, 1963, pp.316–317).

[26] Einstein's view of "light quanta" (*Lichtquanta*) was rejected for over a decade by physicists like Planck, Bohr and Millikan ("a reckless hypothesis"). Textbooks fundamentally distort or ignore the historical context according to Niaz *et al.* (2010), who examined over 100 introductory university textbooks.

[27] Both Young and Fresnel wave theories were initially rejected by scientists (Holton and Brush, 2001, p. 348), although many introductory textbooks falsely imply that they represent crucial experiments.

Table 12.1

Distribution of students' views on the NOS on the first and last days of class.

	Bacon	Popper	Kuhn	Lakatos	Other
First day	4	2	0	0	2
Last day	0	5	2	0	1

First and foremost, they see that there are different views according to the different philosophers as to the answer to this question. Students become aware that the same textual material can be viewed in a variety of ways. On the first and last days of class, students are asked to write about the following philosophy of physics questions: "In your view how does science work? How do theories come about and how do new theories take the place of older ones?"

Taking the disagreement of learning theory into consideration, at least about half of the students who had a view that could be expressed as Popperian on the first day of class could be said to have a coherent view of science. At the end of the course, this changed so that only the three students labeled as "others" had a view of science that could be categorized as "knowledge in pieces." Those students categorized as Baconian had sharpened their viewpoint and now had a clearly coherent view of the *nature of science* (NOS). Kalman (2010) explored the progress of eight students throughout the course. Their views on the NOS on the first and last days of class are found in Table 12.1.

This provides evidence that students have met one of the goals of the course (a major step in producing conceptual change)—a critical analysis of their belief structure.

On the final examination, students are required to examine a scientific theory that they have never seen before from the point of view of all philosophers studied in the course. Almost all of the students were able to answer the question at a satisfactory level. This provides evidence that students have met another of the major goals of the course—the ability to analyze textual material from different points of view.

12.6.2.2 Using reflective writing and the hermeneutical circle method

PER has equally shown that students entering and leaving physics courses quite often have their original faulty intuitive ideas of motion, occasionally resembling "Aristotelean" type-views, uninfluenced and intact (McCloskey, 1983; and Schulz, 2019).

In the Newtonian model and metaphysics, "force" changes the state of motion (velocity) and is a cause of motion (external) but not an entity *of* the body (internal). Yet for many students, force is in the body, and it is proportional to velocity. Such a student conception is not only something learned from childhood and "common sense" but also accepted by pre-Galilean natural philosophers. Galileo himself began with this erroneous notion as found in his first (ca.1590) unpublished book *De Motu* [On Motion]. The key change to the modern conception of force occurred with Galileo's discovery of the law of inertia (Kalman, 2009).

If students believe that force is in the body and proportional to velocity, how can we enable them to change to an incommensurate theory where force is external and proportional to acceleration? According to whether students' views are coherent (alternate framework idea) or fragmented (p-prims), would impact the kind of strategies used in the educational process. Feyerabend (1993) in his *principle of counter induction* states that a change in theories—the cognitive shift from one to another—can only occur when one (paradigmatic) theory is contrasted with another: in other words, when an *alternative* theory is available for comparison. (This is the basis of historical controversies, as we've seen). By analogy, students must be given the opportunity to consider multiple competing theoretical models – whether generated by students through the modeling process (Lattery, 2016) or occurring in comparison with the theories proposed by different student groups (Kalman and Rohar, 2010; and Kalman *et al.*, 2015)—in order to change their naïve conceptions to canonical ones (i.e., a switch of their epistemologies).

How can this be effectively achieved? The literature has suggested various conceptual change strategies with more or less success. Instead, we focus on the use of the *Reflective Writing* (RW) tool and the *hermeneutic circle* method. The RW tool is a metacognitive activity that prompts students to examine textual material *before* coming to the classroom or laboratory, in the manner of "turning" in a hermeneutical circle (El-Helou and Kalman, 2018).

The hermeneutic circle is based on the idea that to understand a text, one must understand its parts and the parts are understood only through their relation to the whole text itself. With every 'turn of the circle' (movement back and forth), a reader improves his or her understanding of the textual content. Another problem with current textbooks is that reading and *interpreting* them (or notes from a lecture) present new kinds of obstacles to learning: written from the viewpoint of the expert, their technical vocabulary and abstract language can seem just as strange as learning a foreign language, which tends to alienate instead of encouraging engagement.

Gadamer (2004) introduced the modern theory known as *philosophical hermeneutics* in 1960. He viewed the hermeneutic circle as the intersection of two language "horizons:" the horizon of the text and that of the reader. The latter has dynamic boundaries that are determined; they evolve with the knowledge, lived experience and skill sets of the reader. But the first horizon of "the text we are trying to *interpret* also has its horizon: a limit to all those meanings to which a text of this sort, employing a language of this sort, could give rise" (Eger, 1993, p. 14). If the student's and text's language horizons do not *overlap*, there is no way for the projections of the student to fall within the realm of the text's potential meanings; thus, the attempt to reach understanding fails (e.g., students' misconceptions).

RW is designed to trigger questioning and enhance connections: "Reflective writing involves reading a section of a textbook, trying to understand it as best as possible, and then using a form of writing called *freewriting* as a tool to self-dialogue about concepts found in the section" (Kalman, 2017, p. 14). This clarifies their own thinking and helps expose their epistemology. It liberates students to get to and explore ideas, where they do not censure their thoughts as they might in open class discussions. True understanding occurs with the "fusion of horizons" as a result of the reader (the student) being engaged in the hermeneutic circle of the textbook language and the instructor's discourse. Gadamer repeatedly emphasized the central role that questioning plays in the back-and-forth process of the circle.

In Kalman and Rohar (2010), students followed RW with a Conceptual Conflict Collaborate Group (CG) exercise and an argumentative essay (critique) using conflicting ideas of Aristotle, Galileo, Newton, and others. In the CG for each exercise, students are asked to discuss (for a fixed time limit) a demonstration or qualitative problem. The lesson impresses the student that there are at least two ways of looking at the problem. Having two groups with different concepts report to the class produces the desired conceptual conflict implicitly using Feyerabend's *principle of counter induction*. Then, representatives of each group debate the issue between themselves. Afterwards, the rest of the class was invited to present questions to these representatives. To underscore that two conflicting concepts have been presented, the class is asked to vote on which concept resolves the demonstration or qualitative problem. (Voting is essential because students, due to cognitive dissonance, often misinterpret what they hear or read.) Due to the vote, students are anxious to find out which point of view is correct and remain very engaged. The instructor resolves the conflict using demonstrations.

The "critique activity" was introduced to promote a critical examination of the alternatives produced in the collaborative group exercise. Students have to produce as many possible arguments that favor all of the conceptual ideas raised in class and then indicate which viewpoint is in accord with the experimental evidence. The critiques are designed to encourage the students to undergo a "critical discussion to decide which natural interpretations can be kept and which must be replaced" (Nelson, 1994).

All studies discussed helped students move toward a conception of physics and scientific knowledge as a highly ordered, coherent, knowledge structure that contains a set of interrelated ideas. Also, the nature of theory change is complex and requires argumentation. Finally, they realize that textbook content knowledge, experiments and demonstrations can be viewed from different philosophical perspectives.

12.7 CONCLUSION

This chapter has surveyed how philosophy has played a major role in the development of physics, both as philosophy-*in* and -*of* physics. The academic sub-discipline PoP was also inspected—a unique presentation since that sub-discipline has generally been overlooked in physics education. Philosophy as a subject is unavoidable where experiments, laws and theories are to be interpreted and creative proposals are imagined. Hence, the philosophy of physics is inherent to the practice, advancement and historical progression of physics, and it cannot and should not be avoided by instructors at secondary and undergraduate levels. Research pertaining to learning theories and instructional strategies was presented that linked better student comprehension, inclusive of critical thinking, directly to PoP themes. A limitation of this work is acknowledged in so far as the philosophy emphasis has been restricted to the metaphysics and epistemology of physics. Future research should include the aesthetics, ethics and sociology of physics with respect to educational aims.

REFERENCES

Adler, C. G., Am. J. Phys. **55**(8), 739–743 (1987).
Albert, D. Z., see https:// www.britannica.com/topic/philosophy-of-physics for Philosophy of physics. Encyclopedia Britannica (2018) (last accessed December 17, 2020).
Aristotle, *Physics* (Oxford University Press, Oxford, 1996).
Aristotle, *The Metaphysics* (Penguin, London, 1998).
Arons, A., *Teaching Introductory Physics* (Wiley, 1997).
Bächtold, M. and Guedj, M., *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. Matthews (Springer, 2014), pp. 211–245.
Bacon, F., in *the Works (Vol. VIII)*, translated by J. Spedding *et al.* (Longman & Co., Boston, 1863), pp. 1857–1874.
Barnes, J., *Early Greek Philosophy* (Penguin, 1987).
Batterman, R. (ed.), *The Oxford Handbook of the Philosophy of Physics* (Oxford University Press, 2013).
Bauer, H. H., *Scientific Literacy and the Myth of the Scientific Method* (University of Illinois Press, Chicago, 1992).
Becker, A., *What is Real? The Unfinished Quest for the Meaning of Quantum Physics* (Basic Books, New York, 2018).
Bell, J. S., *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed. (Cambridge University Press, Cambridge, 2004).
Belot, G. and Earman, J., *From Physics to Philosophy*, edited by J. Butterfield and C. Pagonis (Cambridge University Press, Cambridge, 1999), pp. 166–186.
Besson, U., *International Handbook*, edited by M. Matthews (Springer, 2014), pp. 245–283.
Bigaj, T. and Wüthrich, C., (eds.) *Metaphysics in Contemporary Physics* (Brill, 2015).
Born, M., *Einstein's Theory of Relativity* (Dover, 1962).
Bowler, P. J. and Morus, I. R., *Making Modern Science. A Historical Survey* (University of Chicago Press, Chicago, 2005).
Braga, M. *et al.*, Sci. Educ. **21**(6), 921–934 (2012).
Bunge, M., *Philosophy of Physics* (Dordrecht, 1973).
Bunge, M., *Finding Philosophy in Social Science* (Yale, 1996).
Bunge, M., Sci. Educ. **21**(10), 1601–1610 (2012).
Burtt, E. A., *The Metaphysical Foundations of Modern Physical Science*, revised ed. (Doubleday, 1954).
Butterfield, J. and Earman, J. (eds.), *Philosophy of Physics* (North-Holland Press, Amsterdam, 2007).
Callender, C. and Huggett, N. (eds.), *Physics Meets Philosophy at the Planck Scale: Contemporary Theories of Quantum Gravity* (Cambridge University Press, Cambridge, 2001).
Cartwright, N., *How the Laws of Physics Lie* (Clarendon, Oxford, 1983).
Chi, M. T. H., *International Handbook of Research on Conceptual Change* (Routledge, 2013), pp. 47–70 .
Clagett, M., *Greek Science in Antiquity* (Collier, 1963).
Coelho, R. L., Sci. Educ. **19**(1), 91–113 (2010).
Coelho, R. L., Sci. Educ. **21**(9), 1337–1356 (2012).
Coelho, R. L., Sci. Educ. **22**(5), 1043–1068 (2013).
Cordero, A. (ed.), *Philosophers Look at Quantum Mechanics* (Springer, 2019).
Crull, E. M., Anal. Rev. **73**(4), 771–784 (2013).
Cushing, J. T., Interchange **20**(2), 54–59 (1989).
Cushing, J. T., *Philosophical Concepts in Physics. The Historical Relation Between Philosophy and Scientific Theories* (Cambridge University Press, Cambridge, 1998).
Cutnell, J. D. *et al.*, *Physics*, 11th ed. (Wiley, 2018).
Dainton, B., *Time and Space*, 2nd ed. (McGill-Queens Press, Montreal, 2010).
De Haro, S., Found. Sci. **25**, 297–314 (2020).

d'Espagnat, B., *On Philosophy and Physics* (Princeton University Press, Princeton, 2006).
diSessa, A. A., Cognition Instruction **10**(2–3), 105–225 (1993).
diSessa, A. A., *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, 2008), pp. 35–60.
DiVincenzo, D. and Fuchs, C., Phys. Today **72**(2), 50–51 (2019).
Dolnick, E., *The Clockwork Universe. Isaac Newton, the Royal Society and the Birth of the Modern World* (Harper, 2011).
Eger, M., Am. J. Phys. **40**, 404–415 (1972).
Eger, M., Sci. Educ. **2**(1), 1–29 (1993).
Einstein, A., *Relativity. The Special and the General Theory* (Methuen, 1920/1954).
Einstein, A., *Ideas and Opinions*, edited by A. Einstein (Crown, 1934/1954), pp. 276–285.
Einstein, A. and Infeld, L., *The Evolution of Physics. The Growth of Ideas From Early Concepts to Relativity and Quanta* (Simon and Schuster, 1938).
El-Helou, J. and Kalman, C. S., Phys. Teach. **56**, 88–91 (2018).
Feyerabend, P., *Against Method*, 3rd ed. (Verso, 1993).
Feynman, R. P., *Six-not-so-easy pieces: Einstein's relativity, symmetry and space-time* (New York, 1997).
Franck, P., Am. J. Phys. **15**, 202 (1947).
Gadamer, H.-G., *Truth and Method* (Continuum, 2004).
Garritz, A., Sci. Educ. **22**(7), 1787–1807 (2013).
Gatti, H., *Giordano Bruno and Renaissance Science* (Cornell, 1999).
Giancoli, D. C., *Physics: Principles with Applications*, 4th ed. (Pearson, 1995).
Gil, D. and Solbes, J., Int. J. Sci. Educ. **15**(3), 255–260 (1993).
Greca, I. M. and Freire, O., *International Handbook*, edited by M. Matthews (Springer, 2014), pp. 183–210.
Hadzigeorgiou, Y. and Schulz, R. M., Sci. Educ. **23**, 1963–2006 (2014).
Hadzigeorgiou, Y. and Schulz, R. M., Educ. Sci. 7, 84 (2017).
Hadzigeorgiou, Y. and Schulz, R. M., Front. Educ. **4**(38), 1–10 (2019).
Halliday, D. *et al.*, *Fundamentals of Physics*, 10th ed. (Wiley, 2013).
Harré, R. *et al.*, Metascience **19**, 349–371 (2010).
Hawking, S. and Mlodinov, L., *The Grand Design* (Bantam, 2010).
Hecht, E., Phys. Teach. **44**, 40–45 (2006).
Hecht, E., Eur. J. Phys. **37**, 065804 (2016).
Heisenberg, W., *Physics and Philosophy. The Revolution in Modern Science* (Harper, 1958/2007).
Hesse, M. B., *Forces and Fields. The Concept of Action at a Distance in the History of Physics* (Greenwood, 1962).
Hoefer, C., Br. J. Philos. Sci. **49**, 451–467 (1998).
Holton, G. and Brush, S. G., *Physics, the Human Adventure. From Copernicus to Einstein and Beyond* (Rutgers, 2001).
Höttecke, D. and Silva, C. C., Sci. Educ. **20**, 293–316 (2011).
Jeans, J., *Physics and Philosophy* (Dover, 1943).
Kalman, C., Sci. Educ. **11**, 83–94 (2002).
Kalman, C., Sci. Educ. **18**, 25–31 (2009).
Kalman, C., Sci. Educ. **19**(2), 147–163 (2010).
Kalman, C., *Successful Science and Engineering Teaching in Colleges and Universities*, 2nd ed. (Charlotte, 2017).
Kalman, C. and Lattery, M. (eds.), Front. Psychol. **10**, 2861 (2019).
Kalman, C. S. and Rohar, S., Phys. Rev. Spec. Top.: Phys. Educ. Res. **6**(2), 020111 (2010).
Kalman, C. S. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **11**(2), 1–17 (2015).
Knight, R. D., *Physics for Scientists and Engineers* (Pearson, 2017).
Knox, E. and Wilson, A. (eds.), *The Routledge Companion to Philosophy of Physics* (Routledge, 2021).
Koertge, N., *A House Built on Sand. Exposing Postmodernist Myths About Science* (Oxford, 1998).
Kosso, P., *Appearance and Reality: An Introduction to the Philosophy of Physics* (Oxford, 1998).
Koyré, A., *From the Closed World to the Infinite Universe* (Hopkins, 1957).
Kragh, H., *International Handbook of Research*, edited by M. Matthews (Springer, 2014), pp. 643–665.
Kuhlmann, M. and Pietsch, W., J. Gen. Philos. Sci. **43**, 209–214 (2012).
Kuhn, T., *The Copernican Revolution* (Harvard, Cambridge, 1957).
Kuhn, T., *The Structure of Scientific Revolutions*, 2nd ed. (Chicago University Press, Chicago, 1970).
Kuhn, T., *The Essential Tension* (Chicago University Press, Chicago, 1977), pp. 127–146.
Ladyman, J., *Understanding Philosophy of Science* (Routledge, 2002).
Lam, V. and Esfeld, M., J. Gen. Philos. Sci. **43**, 243–258 (2012).
Lange, M., *An Introduction to the Philosophy of Physics* (Blackwell, 2002).
Lattery, M. J., *Deep Learning in Introductory Physics: Exploring Studies of Model- Based Reasoning* (Charlotte, 2016).
Laudan, L., *Philosophy of Science. The Central Issues*, edited by M. Curd and J. Cover (Norton, New York, 1981/1998), pp. 1114–1135.
Levrini, O., Sci. Educ. **11**(3), 263–278 (2002).
Levrini, O., *International Handbook of Research*, edited by M. Matthews (Springer, 2014), pp. 157–182.
Lucretius, T., *On the Nature of the Universe* (Penguin, 1951).

Machamer, P., *Scientific Controversies. Philosophical and Historical Perspectives*, edited by P. Machamer *et al.* (Oxford University Press, Oxford, 2000), pp. 1–17.
Margenau, H., *Physics and Philosophy: Selected Essays* (Dordrecht, 1978).
Marquit, E., Am. J. Phys. **46**(8), 784–789 (1978).
Marzano, R. J., Theory Pract. **32**, 154–160 (1993).
Matthews, M. R., *Science Teaching. The Contribution of History and Philosophy of Science*, 20th revised ed. (Routledge, 2015).
Maudlin, T., *The Metaphysics Within Physics* (Oxford University Press, Oxford, 2007).
Maudlin, T., *Philosophy of Physics. Space and Time* (Princeton University Press, Princeton, 2012).
Maudlin, T., *Philosophy of Physics. Quantum Theory* (Princeton University Press, Princeton, 2019).
McCloskey, M., *Mental Models*, edited by D. Gentner and A. L. Stevens (New York: Psychology Press, Hillsdale, 1983), pp. 299–324.
Muller, F. A., Found. Sci., **28**(1), 477–488 (2023).
Nelson, C., *Collaborative Learning: Underlying Processes and Effective Techniques*, edited by K. Bosworth and S. J. Hamilton (Jossey-Bass, 1994), pp. 45–48.
Niaz, M. *et al.*, Sci. Educ. **94**, 903–931 (2010).
Norris, C., Found. Sci. **5**, 3–45 (2000).
Norsen, T., *Foundations of Quantum Mechanics* (Springer, 2017).
Ohanian, H. C., *Einstein's Mistakes. The Human Failings of Genius* (Norton, 2008).
Pais, A., *Niels Bohr's Times, in Physics, Philosophy and Polity* (Clarendon, Oxford, 1991).
Plato, *Plato's Theory of Knowledge. The Theaetetus and Sophist* (Dover, 2003).
Popper, K., *Conjectures and Refutations* (Routledge, 1963).
Posner, G. *et al.*, Sci. Educ. **66**, 211–227 (1982).
Price, H., *Time and Matter* (World Scientific, 2006), pp. 209–224.
Richards, R. J., *The Romantic Conception of Life. Science and Philosophy in the Age of Goethe* (Chicago University Press, Chicago, 2002).
Rickles, D., *The Philosophy of Physics* (Polity Press, Cambridge, 2016).
Rickles, D. (ed.), *The Ashgate Companion to Contemporary Philosophy of Physics* (Ashgate, Farnham, 2008).
Roche, J., Eur. J. Phys. **26**, 225–242 (2005).
Romero, G. E., *Mario Bunge: A Centenary Festschrift*, edited by M. Matthews (Springer, 2019), pp. 289–302.
Rovelli, C., Found. Phys. **48**, 481–491 (2018).
Schulz, R. M., *International Handbook of Research*, edited by M. R. Matthews (Springer, 2014a), pp. 1259–1316.
Schulz, R. M., *Rethinking Science Education. Philosophical Perspectives* (IAP, Charlotte, 2014b).
Schulz, R. M., Sci. Educ. **28**, 1273–1278 (2019).
Seeger, R. J., Am. J. Phys. **28**, 384–393 (1960).
Sklar, L., *Philosophy of Physics* (Oxford University Press, Oxford, 1992).
Sklar, L., Br. J. Philos. Sci. **51**, 729–742 (2000).
Sklar, L., *Philosophy and the Foundations of Dynamics* (Cambridge University Press, Cambridge, 2013).
Smart, J. J. C. (ed.), *Problems of Space and Time* (Macmillan, New York, 1964).
Smolin, L., *The Trouble with Physics. The Rise of String Theory, the Fall of a Science and What Comes Next* (Mariner, 2006).
Torretti, R., *The Philosophy of Physics* (Cambridge University Press, Cambridge, 1999).
Vosniadou, S. *et al.*, *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, 2008).
Watson, P., *The German Genius. Europe's Third Renaissance, the Second Scientific Revolution, and the Twentieth Century* (Harper, 2010).
Weinberg, S., *Dreams of a Final Theory* (Pantheon, 1992).
Westfall, R. S., *The Construction of Modern Science: Mechanism and Mechanics* (Cambridge University Press, Cambridge, 1971).
Young, H. and Adams, P., *Sears & Zemansky's College Physics*, 11th ed. (Pearson, 2020).
Zinkernagel, H., Theoria **26**(2), 215–241 (2011).

CHAPTER

13

NEW METHODOLOGICAL APPROACHES TOWARD IMPLEMENTING HPS IN PHYSICS EDUCATION: THE LANDSCAPE OF PHYSICS EDUCATION

Elizabeth Mary Cavicchi, Hillary Diane Andales, and Riley S. Moeykens

Cavicchi, E. M., Andales, H. D., and Moeykens, R. S., "New methodological approaches toward implementing HPS in physics education: The landscape of physics education," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 13-1–13-32.

13.1 INTRODUCTION

Physics is a human enterprise entangled in a web of historical, philosophical, social, and cultural contexts. Yet in most physics classrooms, this humanity is hidden from students. Physics is decontextualized. Under decontextualized teaching, students encounter a false kind of physics—one that is isolat*ed* and isolat*ing*. When physics loses its context, students lose the means to deepen their understanding and relationships with physics.

This chapter highlights contextualized teaching methods that reintegrate physics into its web of context, emphasizing the history and philosophy of science (HPS). Because context introduces humanity, complexity, identity, and voice to an otherwise sterile learning experience, contextualized methods welcome all students, invite their democratic participation and flatten classroom hierarchies. To represent the teaching methods in our chapter, we present a metaphorical Landscape of Physics Education (Fig. 13.1). In our landscape, each teaching method occupies its own space. We broadly characterize these

https://doi.org/10.1063/9780735425514_013

The International Handbook of Physics Education Research: Special Topics, Mehmet Fatih Taşar and Paula R. L. Heron

FIG. 13.1

The Landscape of Physics Education: Decontextualized methods live at high elevations where the learning experience is exhausting and suffocating. The ecosystem is barren, reflecting the sterility of decontextualized physics instruction. When descending along any path, contextualized methods add meaning; water represents that context. Exhibiting many forms—streams, waterfalls, puddles, rivers, and marshes—water acts as a nourishing agent, enriching the learning experience, and as an eroding agent, abrading the foundations of decontextualized methods. At the lowest elevations, physics is fully contextualized. Students learn in democratic spaces where their identities, voices, and agencies are celebrated. Notice how students—our main characters—interact with the landscape at each elevation. Design by Hillary Diane Andales.

methods as *contextualizing pedagogies,* which include historical experiments, science stories, historical narratives, and philosophy, with openings for feminist science and indigenous knowledge.

Related literature reviews apply a landscape model to educational research: “the landscape of HPS educational practices …is manifold” (Henke and Höttecke, 2015, p. 350); “evolving landscape” in nature of science (NOS) assessments (Abd-El-Khalick, 2014, p. 621); and the “pedagogical landscape” regarding student identities from underrepresented groups (Rader, 2020, p. 568). This chapter’s landscape shares those features: diverse landforms, ongoing change, and interaction between landforms and student identities.

Our landscape uses two central metaphors: altitude and water. The landscape schematic has three elevation levels: high, middle, and low. At high elevations, physics is decontextualized. The formidable peak is only accessible to students who are already equipped with the right resources and "climbing skills." Just as climbing mountains are exhausting, for many students the experience of learning physics without context is intimidating. Decontextualization propagates an image of science as elitist, exclusive, and dismissive of personal contributions (Adúriz-Bravo and Pujalte, 2020, p. 216).

At middle elevations, educators aim for methods that make physics more understandable and accessible to diverse students. Isolated puddles of context are introduced. Here, the decontextualized curriculum is interspersed with science stories, historical experiments, indigenous knowledge, feminist science, and so on.

Finally, at the lowest elevations, where the ecosystem is diverse and accessible, physics is fully contextualized. Here, teachers prioritize both content *and* context. Students' voices and identities are celebrated. At this elevation, water, the other main metaphor, is also most prominent.

Water represents context in education. It functions as a nourishing agent, enriching learners' experiences. It also acts as an erosive agent, eroding decontextualized methods. Water comes in many forms—streams, waterfalls, puddles, rivers, and marshes—much like how contextualized methods of instruction come in different forms. In this chapter, we argue for teaching methods that function like water at every elevation.

In addition, we encourage educators, institutions and students to experiment with doing education at low elevations. This educational effort inverts the conventional thinking pervasive in academia that values high-level knowledge and devalues whatever is low. In our landscape model, education is most productive at the low elevations. While at high elevations, students are actively discouraged from questioning high-level knowledge, at the lowest elevations, students are surrounded by immersive opportunities to observe and learn. The curriculum is open and co-created among students, teachers, the environment, physical phenomena, history, philosophy and the wider world. Anyone can participate, wherever they are in the world and in life circumstances. Students are empowered to explore and act on possibilities and curiosity. Higher elevations are inhospitable; lower elevations are more suitable for living and learning.

A novice physics student is the "main character" in this chapter's landscape. This chapter includes students' perspectives and voices across the research examples and themes selected for discussion. By centering the young physics student's perspective in this landscape, we give students their rightful place as the focus in education. We also admit that this landscape is subjective. One student might see the marsh as better than the mangrove; another might think otherwise. This reflects how different students respond to different HPS or related contextualizing methods. Nevertheless, decontextualized physics is difficult for *all* students: though some may have a propensity to climb its peaks, that climb will remain difficult.

This chapter takes readers on themed “tours” that visit areas where “water” is active in contextualizing the Landscape of Physics Education. As our literature survey demonstrates, there are many methods for introducing the contextualizing action of water to physics education. Innumerable tours are possible. Of those possibilities, we describe three tours: historical experiments and instruments, historical narratives and story-telling, and openings to feminism and indigenous knowledge.

Each tour commences at the high elevations that dominate physics education today, where instruction is hierarchically structured and transmitted, with students in passive roles. Proceeding to mid-elevations, tours move along places carved by and pooled with water, where the context becomes stable and students are more actively involved. “Stops” on these tours correspond to specific examples relating to that tour’s theme. At low elevations, tours overlap in collaborative experiences.

13.2 CHAPTER ORIENTING ASPECTS

13.2.1 Chapter author perspectives

The American Physical Society (APS) urges efforts toward “greater diversity” in physics education and careers (APS, 2018) and for “enabling full participation of women in physics” (APS 2015). In support of these aims, physics education researchers integrate analyses and practices relating to experiences and context of underrepresented groups and women, including critical race theory (CRT; Nissen *et al.*, 2021; and Rodriguez *et al.*, 2022), feminist standpoint theory (Rodriguez *et al.*, 2022) and LGBT + identity (Barthelemy *et al.*, 2022). Respecting feminism’s finding that all knowledge reflects its specifics and human creators, and that knowledge’s objectivity is enhanced by including multiple human perspectives, physics education researchers provide statements of their own “positionality” in that process (Avraamidou, 2020; Nissen *et al.*, 2021; and Robertson and Hairston, 2022). As chapter authors, we support that practice by noting our perspectives in the physics education landscape.

In the high elevation physics of her MIT undergraduate studies and subsequent teaching, Elizabeth Cavicchi found herself disempowered. This impasse changed through her involvement with alternative and democratic understandings of education (Dewey, 1916, 1934; Friere, 1970, 2000; Morrison, 1995; Hawkins, 2002; and Duckworth, 2006). Presently teaching at MIT’s Edgerton Center, Cavicchi encourages low elevation experiences where all participants, including the teacher, are in the unknown and collaboratively create investigations (Cavicchi, 1999, 2011, 2014, 2018, 2021a, and 2021b).

For student-author Riley Moeykens, transformation into a physicist was not one of deep understanding of the field but rather a transformation of one’s very being. The stereotype of a physicist was deeply ingrained as such: detached, unemotional, individualistic, and even masculine. Thus, Riley Moeykens’ desire to engage with physics in the physics classroom only seemed possible through the transformation of her own character into an unachievable stereotype. Under the guidance of Elizabeth Cavicchi, Moeykens began to find her own personal value as a physics student and began to recognize that physics is not isolated from humanity. She realized that true physics education encourages student voice and is accepting of all identities and the role they play in shaping one's experience with physics.

To student-author Hillary Andales, physics previously represented wonder and escape—an objective endeavor insulated from worldly squabbles. However, in late high school, water began eroding her decontextualized understanding. She learned about how Einstein's science was spurned by his German peers because he was Jewish, how Manhattan Project physicists translated math equations to mass destruction, how a complex chain of subjective negotiations co-creates seemingly objective brain images (Dumit, 2004), and so on. Given how a solid understanding of the nature of science is becoming more imperative in our increasingly science-driven world (Dagher, 2020; and McComas, 2020a), Hillary Andales seeks to integrate context in her science communication work.

13.3 WIDELY DOCUMENTED TRENDS

Among research that, like water, gradually erodes and levels the Landscape of Physics Education, we find two areas that are widely documented and researched: "constructivism" and "nature of science." These areas are considered by many educational researchers as matters "already known." We briefly orient the reader to these areas.

13.3.1 Constructivism

Educational work that addresses the contexts of science through historical, philosophical, and social dimensions often addresses learners' context and agency (de Castro and de Carvalho, 1995; Aikenhead, 1996; Cobern, 1996; and Klassen, 2006). Educational research that treats students as active participants with their own context and agency, as opposed to passive receptors, is associated with constructivism. In a constructivist introductory physics course, one engineering student, exhilarated over grappling with the unknown, contrasted that experience with conventional instruction:

> it was just really about learning. […] It is about the journey and the question. It wasn't about absolute right or wrong (Radoff *et al.*, 2019).

The finding that learners are not passive receivers but rather construct their own meanings is substantiated by a research basis encompassing craft traditions, Dewey's educational analyses (1916, 1933), research by Piaget (1952, 1965) with Inhelder (1958); and Vygostky (1962) and the liberatory work of Freire (1970). Applying this research involves students in dynamic interaction with the environment and community (Dewey, 1934; Wong and Pugh, 2001; Neff and Helwig, 2002; and Johnston, 2014). In contrast, hierarchical instruction inhibits student interactions, functioning to impose colonization on them (Larochelle *et al.*, 1998; Whitten, 2012; and Robertson and Hairston, 2022).

In creating spaces where experiences evolve, constructivist teachers are co-learners, listeners, researchers and facilitators (Lincoln, 1995; Barton, 1997; Kubli, 2005; and Duckworth, 2006) with students as partners (Sohr *et al.*, 2020). Constructivist practices include discussion, brainstorming, experiment, presentation, readings, lectures, and more (Stinner, 1995; Allchin, 1999; Klassen, 2006; and Adúriz-Bravo, 2014). Learners raise contexts including history, philosophy, social justice, every day and indigenous life (Cobern, 1996; Klassen, 2006; Whitten, 2012; and Zidny *et al.*, 2020). Developing agency, autonomy, and democratic skills, students are not confined by power structures (Larochelle, 1998; Neff and Helwig, 2002; Sriprakash, 2010; and Philip and Gupta, 2020).

Critiques of constructivist approaches reflect tensions regarding academic outcomes. For example, critics note that learners' constructed understanding may conflict with established science (Matthews, 1994; and Seatter, 2003). Whereas conventional instructors mark students' incorrect answers as wrong (Wong *et al.*, 2001), in a constructivist instruction, teachers provide counterexamples; students then evolve collectively (de Castro and de Carvalho, 1995). Impasses for constructivism arise when students' advancement depends on correct answers, teachers' status rests on exam outcomes, or reviewers uphold primacy for formalized science (Osborne, 1996).

13.3.2 Nature of science

"Nature of science" (NOS) encompasses a large body of educational theory, practices and associated research that integrates human and process dimensions of doing science into science education. Broadly, it aims to improve students' understanding of both science content and process (Monk *et al.*, 1997; Teixeira *et al.*, 2012; and Bell *et al.*, 2020). The "consensus" view of NOS prescribes teaching the following concepts (McComas, 2020b):

- Evidence provides a basis for natural laws and scientific theories.
- Science partakes in creativity, bias and culture.
- Science findings are tentative and limited.

Where students grapple with genuine issues while being empowered to develop their own tentative ideas and collaboratively apply evidence critically, they enact "science that is inclusive and democratic" (Dewey, 1916; Morales-Doyle, 2017; and Adúriz-Bravo and Pujalte, 2020, p. 218). By contrast, where students are treated as passive receptors, the curriculum imposes their status as "others;" NOS contextualization functions to "control the other" (Moura *et al.*, 2020, p. 143).

Diverse NOS methods include decontextualized presentations (Park *et al.*, 2019); historical case studies (Allchin, 2011, 2012; and Wong *et al.*, 2014); historical story-telling (Klassen 2006; Metz *et al.*, 2007; Clough, 2011; and Klassen and Klassen, 2014); philosophical debating (Teixeira *et al.*, 2012; Cobern and Loving, 2020; and Dunlop and de Schrijver, 2020); historical experiments (Maurines and Beaufils, 2013; Heering and Höttecke, 2014; Heering and Cavicchi, 2020; and Stefanidou *et al.*, 2020); and investigations on contemporary issues (Wong *et al.*, 2014).

Despite NOS work spanning over a century, being international in participation, and being codified in U.S. and international science education guidelines (Song and Joung, 2014; and Wong *et al.*, 2014), NOS principles remain absent from most physics instruction, including from courses recently taken by coauthors Andales and Moeykens. Although physics education research (PER) has shown how NOS benefits student learning, physics teachers and educational researchers still privilege "the facts" over context (Stadermann and Goedhart, 2021, p. 3).

NOS is also underdeveloped in exposing constructs of power by which those equipped with Western science inflict racism and destruction on indigenous peoples and nature, and on those with nonconforming identities (Chambers and Gillespie, 2000; Gandolfi, 2019; Hansson, 2020; and Ogunniyi, 2020). For instance, NOS does not encompass community-building values such as African "ubuntu" (Ogunniyi, 2020) and "ubanifu" (Semali *et al.*, 2015) as well as African-American "community" (Seiler, 2001)—all of which Western colonization suppressed.

13.4 THE MISSING STUDENT PERSPECTIVE

In viewing the educational research literature from our perspectives as students and former students of physics, chapter authors find notable the absence of students' perspectives from publications integrating HPS with PER. Students often figure in PER as research subjects for trialing curricula and documenting their knowledge. More recently, students have become partners and codesigners in PER that looks to disrupt academic hierarchies. These student-participants transform the questioning, knowledge and process of classroom experiences (Philip and Gupta, 2020; Sohr *et al.*, 2020; and Cardinot *et al.*, 2022). This section comments on two aspects of student experience: Student Voice and Grading. In a future PER landscape, we encourage research that gives voice to student concerns and pain.

13.4.1 Student voice

Where students are acknowledged as actively constructing science understandings, education bears responsibility for providing space to develop and express their voice (Laux, 2018). Authentic student voice relies on a respectful environment, one that accommodates students' risk-taking and is supported by teachers' listening (O'Loughlin, 1995; Cavicchi, 2014; and Jaber *et al.*, 2022). Conventional science education silences student voices, for example, where student contributions are silenced or marked "wrong" (Sriprakash, 2010).

As student voices emerge, power asymmetries in education, science, and society are exposed and questioned. The saying: "You can't teach somebody that don't want to learn" (Seiler, 2001, p. 1006), voiced by Black teen Ed, guided an educational study where Ed, along with other Black male teens, met informally with a White female teacher at an otherwise hierarchically structured high school. Collaboratively, youth initiated and conducted their own physics research, undertaking experimental tests with basketballs and tennis balls (Seiler, 2001).

An immigrant minority undergraduate woman articulated oppressive classroom conditions during conversations with a White male teacher who "just listened." Their co-authored educational research narrates her developing personal agency and authentic voice (Secules *et al.*, 2018). Saying that "life experience and mathematical sides come together," one interviewed graduate student expresses how context matters in his reasoning with quantum mechanics, in a study that attests to the value in education that affirms students' investigations (Dini and Hammer, 2017, p. 13). Coauthor Andales describes this experience of empowering student voices:

> I can think. I critically evaluate science, not just for the facts that it presents but also for how the science was produced, what kinds of power made it. For me as a student, this is empowering (April 2021).

13.4.2 Grading

While grades pervade student experience, grading is noticeably absent in most research involving contextualizing pedagogies. Coauthor Andales calls for educator awareness:

> I think students really just have a concern for Grades. I think Grades should be on the minds of science educators. How high school students balance their understanding of science vs Grades—those two are sometimes on opposite ends of the spectrum (Andales, April 2021).

Grading's inequitable impact on those underrepresented in physics is examined in recent research that regards course design and associated structures, not student deficiencies, as the source of grading disparities (Simmons and Heckler, 2020; Webb *et al.*, 2020; and Paul and Webb, 2022). An extensive analysis of grading in introductory physics demonstrated that underrepresented minorities are penalized inequitably by grading (Paul and Webb, 2022).

Under educational efforts that eliminate grading, assessment focuses on equity, not ranking (Feldman, 2018) and classrooms emphasize mutual respect (Blum and Kohn, 2020). Where grading stresses are replaced with equitable expectations, students experience their own agency in learning, take risks, and collaborate (Feldman, 2018; Gibbs, 2020; and Strommel, 2020). We therefore recommend questioning grading structures in physics education, and trialing alternatives that encourage students to focus on learning, not grades.

13.5 OTHER LITERATURE REVIEWS

The range of methods and studies integrating HPS and related areas with physics and science education is vast. Edited volumes include review chapters on the following: HPS generally (Matthews, 2014) and NOS (McComas, 2020a). Review articles document HPS in physics education (Teixeira *et al.*, 2012), HPS in physics teacher education (Henke and Höttecke, 2015), HPS generally (Monk and Osborne, 1997; Ramadas, 2001; Østergaard *et al.*, 2008; and Heering and Winchester, 2015); NOS (Lederman, 1992; Allchin, 2011; Abd-El-Khalick, 2014; Erduran and Dagher, 2014; Lederman and Lederman, 2014; Dagher and Erduran, 2016; and Lederman and Lederman, 2019); gender, cultures and indigenous science (Aikenhead 1996; Abrams *et al.*, 2014; Traxler *et al.*, 2016; Philip and Gupta, 2020; and Zidny *et al.*, 2020); storytelling and case studies (Klassen, 2006; Metz *et al.*, 2007; Clough, 2011; and Allchin, 2012); student voice (Johnston and Nicholls, 1995; and Laux, 2018); among others.

13.6 TOURS

Informed by the above-noted orienting aspects, we now present three tours through the Landscape of Physics Education (Fig. 13.1):

- historical experiments and instruments;
- contextualization through narratives; and
- feminism and indigenous experience.

Tours 1 and 2 visit educational research projects that make innovative and original contributions while traversing histories and pedagogies having multiple connections to PER practices. Tour 3 confronts longstanding oppressions in physics, physics education, and society, as experienced and addressed in two communities: feminists and indigenous peoples. Inclusive philosophical perspectives and liberatory pedagogies are generative for enacting physics education where each learner's full participation is welcomed, as advocated by recent APS statements (2015, 2018). Work acknowledging oppression and empowering voices conventionally suppressed—including students—is most evident at low elevations in the landscape.

All tours commence at high elevations where physics education is decontextualized, descend to middle elevations where context is increasingly accommodated and arrive at low elevations where learners exchange fluidly with others, physics content and wider contexts.

13.7 TOUR 1: HISTORICAL EXPERIMENTS AND INSTRUMENTS

The stops on this tour report on physics students doing experiments and using instruments, with input from history. At high elevations, where lab work emphasizes pre-specified protocols and results, history introduces learners to ambiguity in what was historically done or understood. Mid-elevations accommodate "historical-investigative (HI) approaches" where students conduct experiments in investigating historically grounded issues while critically evaluating their efforts and the context (Heering and Höttecke, 2014). At low elevations, the open-endedness of students' investigations are commensurate with those of historical experimenting. Unique to this tour is students' direct access to materials, instruments and history. At any elevation, the specific materials that engage students may include authentic artifacts, replicas, commercial instruments, student-built devices, and others.

Before commencing this tour of educational experiences, we review historical research on experiments that distill theoretical insights for education, although not accompanied by classroom trials. These historical studies analyze reasoning from the following:

- phenomena,
- error and uncertainty,
- exploration, and
- still-open questions from historical science.

Examining the historical case of Thomson's e/m experiment, Arons (1982) identified Thomson's productivity in reasoning from phenomena. Applying this finding to physics education, he advocated labs that develop students' physical reasoning, inviting their "guesswork," unlike "narrow and artificial" exercises (p. 19).

Error and uncertainty figure in historical cases researched by Heinicke and Heering (2013; and Heinicke, 2014), where awareness of randomness, and techniques for addressing it emerged as historical investigators dealt with uncertainty at depth—not by canon as decontextualized instruction requires. Proposing that labs be redesigned to involve students in the complexities of data, uncertainty and history, they envision that participating students will form for themselves holistic understandings of physical relationships.

Exploration arises as a crucial means of Goethe's historical color experimenting in Park and Song's (2018) analysis. In projecting how Goethe's exploratory methods might adapt into classroom labs, they foresee students becoming "creative and autonomous" (p. 56) by varying experimental conditions, inferring connections, and initiating dialogue with nature.

Still-open questions from historical science, where science understandings have yet to be determined, provoke Chang's historical research and educational musings (2011). Chang (2011) sees educational potential by involving students in genuine uncertainty, where unresolved experimental questions pose "a live lesson in NOS" (p. 335) having "open-endedness…[that] can only be matched in cutting-edge scientific research" (p. 337). While responding to diverse facets of historical experimenting—such as phenomena, uncertainty, exploration, and still-open questions—these studies concur in advocating students' open and experimental initiatives.

As our tour commences, all stops now address research empirically grounded in classrooms. At this tour's beginning, decontextualized physics is the norm, yet students are provided opportunities for engaging with historical experiments and instruments. Fixed expectations associated with high elevation education abrade historical experimenting's openness.

One strategy that involves bringing entire physics classes to a setting customized for doing historical experimenting is exemplified by the *Einstein was Here* program at the Teylers Museum in the Netherlands (Fig. 13.2). In Lorentz's original lab, high school physics students use replicas

FIG. 13.2
High school physics students experiment with historical electromagnetic instruments at the Teylers' Museum in the Netherlands.

and authentic historical electromagnetic instruments. On arrival, students scoffed at "funny old machines." But after experimenting directly with these artifacts, they appreciated the "incredibly smart" makers and "beautiful" devices (Spek, van der, 2021, p. 259). Students then wrote reports on their experiments in partnership with historical scientists. Given that the visiting classes' local schools may operate at any elevation in the landscape, different reactions surfaced in teacher feedback. Those working at high elevation emphasized students' frustration with the open environment, instead requesting "step-to-step instruction" (p. 261). Teachers who routinely used inquiry methods relayed the students' enthusiasm for the field trip.

At all elevations, experiments originating in history are routinely required by many physics labs, courses and programs. Among these, Millikan's oil drop experiment is notable, being widely implemented as a modern physics lab adapted for instructional settings. Are these instructional labs authentic to its history, experimental practices and educational potential? Investigating many dimensions of instructional labs based on Millikan's oil drop experiment, Heering and Klassen (2010) iteratively developed improvements having greater authenticity. As a baseline, they researched how one German university implemented the lab. This research identified deficiencies: textbook (over-)simplifications fuel unrealistic student expectations and produce feelings of incompetence, problems with commercially produced apparatus, and omission of historical controversy. In response, the researchers revised the lab with historical readings, an online calculator for evaluating data, and a comparison between two commercial apparatuses.

After trialing this revised lab in a Canadian university course, students complained, saying "Looking through a microscope lens for too long makes my head hurt … Don't expect the results to just come" (p. 389). One student said "it was really neat to see the history behind it," while another dismissed history as "not directly necessary" (p. 390). Yielding improved student satisfaction in a subsequent revision, this study highlights multi-dimensional issues arising in efforts to deepen the historical and experimental authenticity of high-elevation physics labs.

Our tour now descends to middle elevations where context matters and education broadens beyond hierarchical information transmission. Historical experiments and instruments are not discrete add-ons to a pre-established curriculum (as in the high-elevation examples above). Instead, they provide openings to the wider student investigation.

An array of historical-investigative (HI) approaches being researched and used in classrooms position physics students as investigators of natural phenomena while concurrently exploring the historical context, experiments, and instruments (Heering and Höttecke, 2014). HI approaches are characterized by: history and philosophy integrated with science; attendance to material, social and cultural aspects of science; NOS education; students' personal explorations of phenomena; and students developing meaning through personal experiences conducting research, critical reflections, reasoning and analyses (p. 1483). Students uncover uncertainty and ambiguity in craft, observation, manipulation and interpretation in the historical accounts and in their efforts. Students self-reflect on contingency in experimenting through their body, skill, experience and mind (Höttecke, 2000).

These functions of HI approaches are illustrated throughout Heering's research account of teaching one unit on electrostatics as a guest in a German high school physics course. At the outset, Heering observed: "it is very hard to overcome the results of a traditional course" (Heering, 2000, p. 370). Students demanded answers to memorize, resisted examining their assumptions and exhibited an entrenched classroom hierarchy where a minority of males dominated while females deferred.

Eventually, this dynamic changed. Genuine bewilderment ensued. Students pointed out inadequacies in their dominant classmates' theories. Meanwhile, the alternative theories developed and tested by the female students gained ground. By that unit's end, Heering reported "not a single student was dominant in the way that I was previously unable to alter…every theory was questioned…or tested experimentally" (pp. 367–368). Students' direct experience with apparatus and phenomena was crucial: historical readings and discussions alone did not provoke the depth of uncertainty, questioning, and interaction that moved students' relationships with physics, learning and each other. With students activated as investigators raising historically informed questions, their classroom elevation shifted from high to moderate and low terrain.

Brazilian high school physics teachers, supported by researchers Batista and Silva (2019), took up the challenge of bringing classes to an unfamiliar place, the cramped school laboratory, for doing HI experiments entailing genuine student inquiry—unprecedented for those teachers and students. Experimental materials were provided via identical (Experimentoteca) kits pre-existing from the 1990s. Brief texts, authored by researchers in consultation with teachers, oriented students to Joseph Black's 18th-century thermal experiments. Each class conducted six lab sessions recorded and analyzed by researchers. Provoking students' rowdy play with flames, the Bunsen burner session did not yield historical inquiry. But after their "second chance" there, the students eventually experimented and collaborated, developing "a feeling of how real science is produced" (p. 1147).

Similarly, historical instrument collections held by some schools catalyze contextualized mid-to-low elevation experiences. History, instruments, and contemporary physics intertwine in classes, internships and thesis projects of physics students in collections at their schools (Wang, 2020; Carchon and Segers, 2021; Cavicchi, 2021b; and Lazos *et al.*, 2021). For example, French high school physics students used the historical Lissajous apparatus to observe interacting frequencies, while masters-level physics students restored instruments and created YouTube videos showing said instruments in action (Khantine-Langlois *et al.*, 2021). After viewing historical instruments at Tsinghua University Science Museum, China, undergraduates developed their interpretations by constructing replicas using workshop tools (Wang, 2020).

Finally, our tour arrives at the low elevations. Here, constructivist practices are explicitly engaged by teachers and students; the academic tensions noted above do not dominate interactions. Less formality is requisite in introducing HI approaches into classroom experimenting. Students access historical and physical materials as peers and teachers encourage curiosity, risk-taking and collaboration.

This educational fluidity is exemplified in two studies in which Brazilian physics teachers, collaborating with educational researchers, engage high school students from low-income families in electrical experiments. In one, the teacher-researcher committed to support learners' agency in constructing personal understandings "without providing final answers" (Silva *et al.*, 2018, p. 331), and invited students to dissect broken appliances during the first lesson. Students' discovery of magnets inside proved inscrutable. One said "plug magnets into wires … to make electricity stronger" (p. 336).

Next, after being posed with a single wire loop, battery, and magnetic compass, the student groups experimented. Many found no role for the compass. When some searched the internet for clues, the teacher countered by encouraging students in working out the phenomena together. After reading Oersted's philosophical interpretation, experimenting resumed. Delving into compass-coil interactions, modeling and debating ideas, they discerned a circular relationship analogous to the "right hand rule" of decontextualized physics. Students' observations and discussions became a source of science knowledge, as in Oersted's historical work.

In the other Brazilian case, teacher-researchers and students expanded the realm of classroom questioning beyond electrical principles through cultural history of science (CHS) perspectives. Despite lacking supplies to replicate the historical Leiden jars they discussed, the students experimented with other electrical materials, and raised issues about scientific funding in historical times and their own. The teacher facilitated an environment of such openness that students articulated how persisting historical inequities strangle their own aspirations, saying:

> When you think about pursuing a scientific career in Brazil, you …cannot only follow your interests… here we have cuts on scholarships… people from [Europe] follow their dreams, they have fewer difficulties, these things matter a lot (Jardim *et al.*, 2021, p. 632).

In further CHS-mediated activities, students developed as "critical negotiators" (p. 634), evaluating societal practices and obstacles integral to science and their personal futures.

Students are initiators of observations and experimenting that they do and of making historical connections in seminars where coauthor Cavicchi involves university students as investigators in science and history (2011, 2014, 2018, 2021a, 2021b). For example, during a class discussion on inverse images, one student gazed out the window and noticed a horizontal bar's "shadow" curved upward, U-like (2018, p. 57). Raising the question "would this shadow invert if viewed from above the bar?", the classmates organized to get themselves higher. It inverted. Later, viewing historical inverted images together, students' thinking about inversion expanded; they realized the role of reflected sunlight in producing the bar's "shadow." One student reflected on how "trusting relationships" with each other and Galileo/history are essential in developing their investigative community, as this example illustrates (2018, p. 60).

Historical experiments and instruments offer educational methods for learners at all elevations to experience materials, phenomena, uncertainty, observation and conjecture. These examples demonstrate the provocative character of those experiences across high, mid and low elevations. Historical research on experimenting, as well as empirical studies of classrooms, support participatory experimenting, such as practiced at low elevations with constructivist and NOS influences. Teacher education students articulate the impact of these methods:

> "Through this experimental procedure we deeply understand refraction." (Stefanidou *et al.*, 2020, p. 9) "I am wondering, how else would we have entered the real world of science if we had not previously opened the door to history?" (p. 6).

13.8 TOUR 2: CONTEXTUALIZATION THROUGH NARRATIVES

We now explore the space of teaching methods that use narratives to reconnect physics content with its context. Narrative-focused methods can be alternative or complementary to historical experiments, the theme of the previous tour. Despite the tactile and immersive experiences of doing historical experiments, teachers may lack training, equipment, preparation time and institutional support for doing them. Narrative methods are accessible through extensive, free and classroom-ready materials. Teachers new to contextualizaton may therefore wish to start with narratives.

In the loosest sense, delivering a narrative is simply "telling someone else that something happened because someone did something" (Metz *et al.*, 2007). Narratives include characters whose actions set in motion a coherent, causally linked sequence of events. Upon including a beginning-middle-end plot structure and a main character who makes a critical choice, a narrative becomes a story (Klassen, 2006; and Klassen and Klassen, 2014). Secondary to these elements are the narrator, narrative appetite (a desire to know what happened), structure, purpose, and receiver (Klassen, 2006, 2009; Metz *et al.*, 2007; and Adúriz-Bravo, 2011, 2014). The terms "narrative" and "story" are used interchangeably in science education literature, and will likewise be used as such in this tour.

Narratives are key to our humanity. A core concept in narrative psychology is the distinction between narrative and paradigmatic/scientific ways of knowing (Bruner, 1986). While these modes are fundamental and irreducible to each other, the narrative mode involves how humans imagine "possible worlds" and ultimately make meaning—cognitive processes that undergird literature, philosophy, everyday thinking and even science itself.

We have an innate narrative way of knowing that supports the effectiveness of stories in science education. By coupling abstract physics content with real-life events and familiar characters, stories engage learners, humanize science, make content memorable and meaningful, stimulate curiosity and encourage reflection on historical and contemporary issues (Arons, 1988; and Martin and Brouwer, 1991). Noddings and Witherell (1991) observe:

> Stories can help us understand by making the abstract concrete and accessible. What is only dimly perceived at the level of principle may become vivid and powerful in the concrete (p. 280).

Science stories seek to teach science content and to humanize the enterprise of science. Drawing from the history of science, historical narratives, also called "cases" or "case studies," typically spotlight a prominent scientist (e.g., Galileo or Nikola Tesla), their journey to discovery, and the historical context within which they made a breakthrough. Though most adhere to historical facts, some invite imagination. Vignettes take historical scientists as main characters addressing fictional events (Roach *et al.*, 1995). Some narratives feature specific contexts, such as climate change, pandemics (Revel Chion and Adúriz-Bravo, 2022), or learners' experiences (Barton and Tan, 2010).

Narrative methods range from self-contained lesson-level stories to more structural unit- and curriculum-level storylines. Stand-alone narrative approaches include anecdotes (Shrigley and Koballa, 1989), short stories (Clough, 2011), case studies (Conant, 1957), and role-plays (Allchin, 2010; and Carvalho and Carvalho, 2002), which can be injected into lessons. In other methods, multiple lessons or entire curricula adopt an overarching narrative (Gorman and Robinson, 1998) or storyline (Stinner, 1995). For instance, Stinner (2006) used the contextualizing problem of "Solar Power in the Pyrenees" for a unit on blackbody radiation.

At high elevations, we find narrative methods that, while offering context, retain authoritative, hierarchical and alienating tones prevalent in traditional science textbooks. Proceeding toward lower elevations, narratives grow in complexity, creativity, and commitment to student voice. Here we find methods such as reflective and mediated science stories (Koliopoulos *et al.*, 2007; and Allchin, 2012), student-led historical excursions (Piliouras *et al.*, 2011; and Paparou, 2021), and meetings with contemporary scientists (Hansson *et al.*, 2019a; and Hadzigeorgiou *et al.*, 2012) which invite more active participation and leadership from learners. Methods here are also conscious and explicit about issues of identity, indigenous experience and politics (Semali *et al.*, 2015; and Moura *et al.*, 2020).

Near the mountain summit, we find the short profiles of famous scientists on the margins of physics textbooks. Though the profiles offer a narrative departure from the bland expositions in the main text, their apocryphal anecdotes and cursory references to history are insufficient. Contextualized narratives must deal with human activities, historical context, and the evolution of understandings about nature (Galili, 2012; and Leone and Rinaudo, 2020).

Case Histories (Conant, 1957) pioneered the introduction of history of science to physics education. It met the above contextualization requirements by using cases such as Boyle's pneumatics and Lavoisier's overthrow of the phlogiston theory, setting the stage for later contextualized works (Rutherford *et al.*, 1970). Years after the release of *Case Histories*, history of science transformed to encompass structural social factors (funding, institutions, gender, class, cultural norms) along with intellectual and biographical contributions—all strikingly absent in *Case Histories*. When evaluated in this new paradigm, the approach taken by *Case Histories* appears narrow and misleading (Allchin, 2011).

At a lower altitude, we find interactive historical vignettes (IHVs), short episodes aiming to put "people back into science" (Roach and Wandersee, 1995). IHVs feature real scientists and real discoveries though narrative elements such as dialogue may be fictional. IHVs are interactive: the teacher pauses narration to ask questions, prompting student reflection about historical, contemporary and personal issues. One college student enjoyed how IHVs "put you in the frame of mind" of the protagonist-scientist (p. 369).

Roleplay (Allchin, 2010; and Carvalho and Carvalho, 2002) pushes learners to acquire research skills, exercise creativity, and practice people and resource management. Nevertheless, co-author Andales observed the drawbacks of roleplay while studying in a specialized science high school in the Philippines. Educational gains (and stress) tended to be concentrated on the learners who already established identities as leaders, creatives, speakers, and honor students. Teachers must therefore design group activities that are more democratic, with opportunities for marginalized students to build strong classroom identity (Secules *et al.*, 2018).

High elevation learners complain that impending tests render history, narrative and roleplay irrelevant. German middle school students objected "Ok, we played along. Now, can we do real physics again?" (Henke and Höttecke, 2015, p. 371). Years of decontextualized instruction conditioned learners to misconceive context as not "real physics." In addressing this issue, teachers must clarify how context helps in internalizing physics content.

Portraying scientists as larger-than-life heroes and science as an object, not a process, high elevation narratives are philosophically incorrect (Milne, 1998), misleading NOS (McComas, 2020a), and alienating. Contrasting strategies for contextualizing narratives offer hints for the lower elevations in our landscape:

> Suspect simplicity. Beware vignettes. Embrace complexity and controversy. Discard romanticized images. Do not inflate genius. Mix celebration with critique. Scrutinize retrospective science-made. Revive science-in-the-making. Explain error without excusing it. And above all respect historical context (Allchin, 2003, p. 347).

Descending, we suddenly find a waterfall whose swift and fierce flow erodes mountains of decontextualized physics. This waterfall includes *Project Physics* (Rutherford *et al.*, 1970) and the Physical Science Study Committee (PSSC; 1960), two contextualized physics courses born out of the United States government's anxiety after Sputnik and became amply supported with government funding (Rudolph, 2002). Meticulously prepared and researched by physicists and teachers, these courses were widely adopted across the United States and internationally. *Project Physics* employed historical contextualization by narrative, while the Physical Science Study Committee (PSSC, 1960) course involved students in constructing apparatus and experimenting with historical, open-ended and everyday contexts. Most textbooks and supplemental materials are digitized and open access (Project Physics, 2010; and PSSC, 2012–2022). Other US physics textbooks (Taylor and Tucker 1941; Glashow, 1994; and Rogers, 1960) centered on historical narratives were less widely distributed.

Having traversed the waterfall to lower elevations, we find a rising trend of research on contextualizing methods. These include: Minnesota Case Study Collection in the United States (Allchin, 2012) with the SHiPS Resource Center (http://shipseducation.net/); The Story Behind the Science (Clough, 2011); and the European HIPST Project (Höttecke, 2012). Various HPS perspectives contextualize the pendulum (Matthews, 2000; and Matthews *et al.*, 2005), a central topic in high school physics. Descriptive, empirical, theoretical and exemplar stories are reviewed by Klassen and Klassen (2014). Metz *et al.* (2007) reviewed door-opening stories, personal narratives, practical work and interrupted storylines, and short stories. The American Institute of Physics (AIP, 2023) produced teaching guides for high school, featuring non-canonical, marginalized physicists: Black nuclear physicist Shirley Ann Jackson, openly transgender neurobiologist Ben Barres, NASA human calculator Katherine Johnson, and Abdus Salam, the first Muslim Physics Nobel Laureate. Dramas and film documentaries (e.g., Melfi, 2016; and Cheney and Shattuck, 2020) raise issues of race, gender, authority and physics in classroom research (Stefanidou, 2016; and Yildirim *et al.*, 2021).

Learners' engagement with stories has potential in activating their learning process through practical, theoretical, social, and affective contexts (Klassen, 2006). For example, a story on Louis Slotin's fatal radiation exposure evokes students' empathy, questions and reflections:

> "Why didn't he throw it out the window?," "How long did it take for the radiation to kill him?," (Klassen, 2009, p. 415)…"The main points … bring a tragic example to light in the wake of doing an experiment on Radiation protection" (p. 417).

Classroom exchange becomes dynamic when the teacher enacts the story. When a teacher, acting as Rutherford, described envisioning an alternative atomic model while eating Christmas pudding, a student interacted directly with Rutherford, saying

> If you hadn't had imagination, you couldn't have thought outside the box…you couldn't have thought "It might be in a different way…" (Hansson *et al.*, 2019b, p. 7).

At a scale spanning all educational levels, Stinner *et al.* (2003) analyze contextual methods employing historical narratives: vignettes, science stories, case studies, narratives, and themes/storylines.

Finally moving to low elevations, students have agency in creating and enacting stories. Twelve-year old students in Greece produce and create their own animated films (Fig. 13.3; Piliouras *et al.*, 2011). One film dramatizes the public mockery toward Tycho Brahe's universe by having tomatoes thrown at it (Fig. 13.4). The children resolve the story by rearranging the tomatoes into heavenly bodies in Brahe's model. About their story-creation experience, two children reflected:

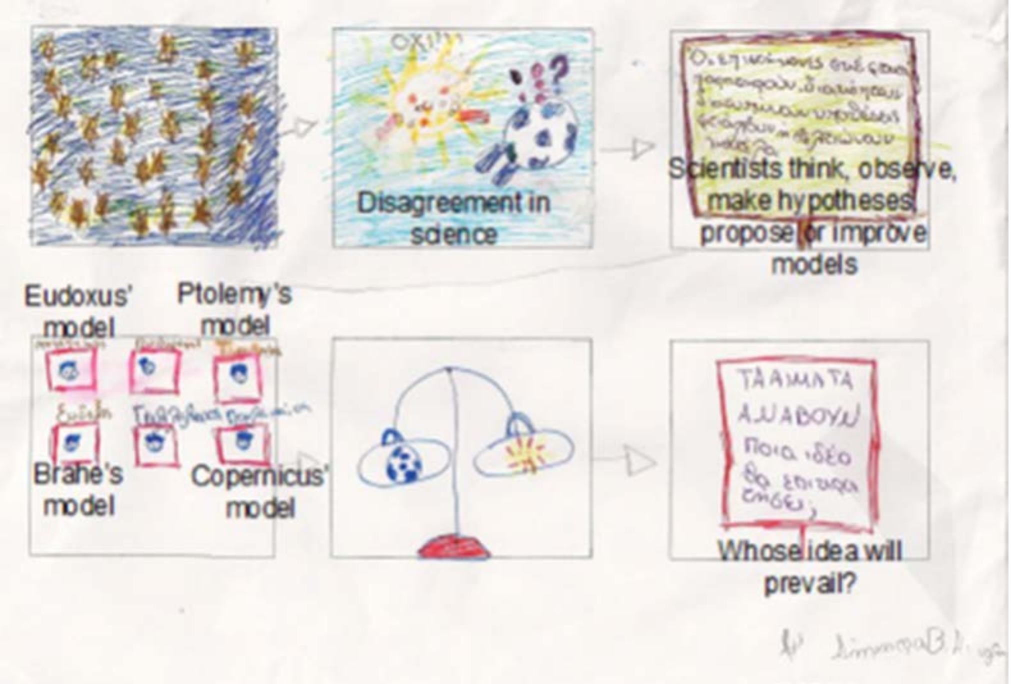


FIG. 13.3
Storyboards developed by schoolchildren in planning their animated films (Piliouras *et al.*, 2011, p. 776).

FIG. 13.4
With the yellow ball as the sun, and planets on wires extending from it, children represent Tycho Brahe's model (Piliouras *et al.*, 2011, p. 776).

> Miltos suggests "We learned about the evolution of science, how things we now learn were formed in the past. (Science) wasn't as we now know it from the beginning; it was shaped in the course of time". And Giorgos adds: "… we learn that something is right and we believe that it will remain forever, but it isn't so" (Piliouras *et al.*, 2011, p. 785).

The youth themselves are the performers and investigators, bringing to light social and scientific issues in the dramas facilitated by Athens physics teacher Paparou. Costarring with Sherlock Holmes, students apply spectroscopic techniques to track down the cause of a historical ailment. Learning is a living drama as "students actively discover" (Paparou, 2021, p. 206). Similarly, Israeli physics teacher education students become intrigued by solving a mystery set by Galileo's lab manuscripts (Schvartzer, 2021).

History can feel distant to learners. To close that distance, Hansson *et al.* (2019a) center contemporary science in teaching NOS, where students meet living and idiosyncratic scientists whom they can emulate as role models (Woods-Townsend *et al.*, 2016). After hearing from contemporary scientists, one undergraduate reflected: "Views were presented [by guest scientists] that differed from my own… I can now approach the issue with better understanding" (Casper and Balgopal, 2020, p. 1578).

Storytelling by contemporary scientists is also powerful. Through collectives such as Skype A Scientist (2019, https://www.skypeascientist.com/) and The Story Collider (2010, https://www.storycollider.org/), scientists share their vulnerabilities and excitement with broader audiences. Through the scientists' narratives, learners see how science is done and how it relates to their lives. Inspired by these resources, the teacher can help guests frame their presentations and help students appreciate connections with classroom content.

However, despite being contemporary, many of the scientists in the above references are only truly relatable to Global North audiences. To engage high school students in Brazil, Cardinot *et al.* (2022) centered Brazil's own science enterprise as a case study, which they could use to learn how science was historically produced. Throughout the research period, students asked probing questions about historical sources and learned about the intersections of science with economic, political and military concerns. Empowered by their student voice, they co-created the curriculum together with the teacher researchers, their questions and personal experiences redirecting the lessons. The free and democratic design of this study exemplifies the constructivist and NOS practices with transformative student involvement throughout.

Guided by the metaphors in our landscape, we explored work that uses narratives for contextual teaching. Gaps include diverse geographical areas and accessibility to teacher resources featuring non-Western and non-canonical figures (Ibn al-Haytham, for example). New narrative mediums offer potential, including video, mobile applications, digital methods (Kouloumtzos and Seroglou, 2011, 2019; and Gentzi and Seroglou, 2019), games and virtual reality.

13.9 TOUR 3: FEMINISM AND INDIGENOUS EXPERIENCE

Whereas previous tours featured well-researched and long-established pedagogies, this final tour focuses on women and indigenous people's relationship with scientific enterprise. These have been matters of concern throughout history and yet have limited influence in today's science education. So they will be focused on for two reasons. First, they center identity and the situatedness of knowledge, which, as the previous tours have shown, are inseparable from context and crucial to making physics accessible. Second, women and indigenous people are underrepresented in physics. Some who study physics report distorting, hiding or altering how they "perform" their identity (Traxler *et al.*, 2016), even how they speak and interact (Ong, 2005, p. 605). How might physics education be changed to welcome students for their inherent selves?

Since PER is only beginning to acknowledge feminist and indigenous educational theory, examples of educational practice are few. This tour thus reviews research from diverse origins, methods, and student voices—from within PER and without. High-elevation "stops" document physics students' adverse experiences relating to their identity. Mid-elevation "stops" present theoretical approaches to feminist and indigenous student experience. Finally, low-elevation "stops" report empirical studies introducing feminist and indigenous perspectives into classrooms.

At high elevations, the dominant "view from nowhere" thinking in science—that is, the idea that disregards the situatedness of knowledge—feigns absolute objectivity and asserts that every person's perspective is equivalent (Haraway, 1988; Cardinot, 2022, p. 3). And yet, historically, women, people of color, indigenous people, and low-income students have been excluded from physics. Students at the intersections of these socially constructed labels (e.g., women of color, low-income indigenous youth) face even greater disadvantages. These identities are subjected to longstanding oppression under dominant cultures, including Western cultures associated with institutionalized physics. This experience of oppression negates the assumption that symmetry pertains between every person's perspective, an assumption underlying "the view from nowhere."

Indigenous peoples are "long-resident, oral cultures peoples" (Snively and Corsiglia, 2001, p. 10). Oppressive colonizing, displacing, and assimilating policies and traumas inflicted by dominating cultures upon indigenous peoples have destroyed lives, cultural works, languages and community practices (Snively and Williams, 2008). Identified as "inheritors and practitioners of unique ways of relating to people and the environment," but now the "most disadvantaged… in the world," (United Nations, 2022), indigenous peoples are distributed worldwide, including Saami in northern Europe, Mayas in Guatemala, and circumpolar Inuit. Neo-indigenous culture refers to non-western culture that, while not of the first peoples of the land, develop long-standing relational understandings (Aikenhead *et al.*, 2007).

Physicist Evelyn Fox Keller (1977) endured continual ridicule, harassment and antagonism during her graduate studies in theoretical physics at Harvard in the mid-1950s. Some physics faculty explicitly challenged her aspiration. She describes such a challenge in the first person, writing: "I was queried about my peculiar ambition to become a theoretical physicist—didn't I know that no woman at Harvard had ever so succeeded" (p. 82). Keenly aware that "she has never known a physicist culturally like her," a physics education researcher, whose identity is both Native American and Chicana, routinely encountered "micro- and macro-aggressions" while navigating her education (Traxler *et al.*, 2016, p. 8). Queer Black cosmologist Chanda Prescod-Weinstein (2021), who reported racist, sexual, and other violations toward her during her physics training, laments how Black physicists are considered "a permanent ontological Other" incapable of objectivity in physics—something which, as an observer phenomenon, should be open to all (Prescod-Weinstein, 2020, p. 424).

These accounts are corroborated by studies of student experience. A British study finds that gender-science relationships, such as those identified by the physicists above, remain unchallenged (Murphy and Whitelegg, 2006). Where "White is the norm" (Archer *et al.*, 2015, p. 201), many Women of Color (WOC) feel alienated; one stated that on entering the lab, "I get the feeling…I'm improperly walking, when I'm in science" (Carlone and Johnson, 2007, p. 1203).

Despite having high grades herself, Black British student Gemma regarded science status as incompatible with the fashionable femininity she favored: "Just clever people are really into Science…their uniform is perfect…or…pointy shoes…" (Archer *et al.*, 2015, p. 216).

Women physicists trained in Muslim majority (MM) countries retrospectively reflect on the congruence between their female Muslim identity, communal goals, and physics. But upon entering non-MM settings, they report conflict between religious and physics identities (Moshfeghyeganeh and Hazari, 2021). Studying in a MM high school, Amina reflected that her female physics teacher "persuaded me to study physics" (Avraamidou, 2020, p. 327). Upon continuing her physics studies in non-MM schools, Amina's intersectional identities as a Muslim and a woman positioned her as "a constant outsider" (p. 314) in physics. Amina was subjected to words spoken "deliberately to hurt" (p. 328).

Similarly, a Māori student describes explicit racism in a science class under a teacher of European descent:

> There is a whole group of us we have our hands up…we know the answer but she'll [science teacher] go to the [Europeans]… before the Māori, like we're second class (Savage *et al.*, 2011, p. 194).

Indigenous students in African communities face disconnects between curriculum content and everyday life (Semali *et al.*, 2015). One African science educator recounts systemic racist propaganda and exclusion during his science education (Ogunniyi, 2020). A student in India contrasts science class with informal learning, saying "Science class is nothing but studying of chapters. There we study, here [informal learning] we can experiment" (Mathai, 2017, p. 21). Entering the science classroom is described as a cultural border crossing, leaving indigenous cultures and entering science culture. When that science subculture is at odds with students' culture, it disrupts their worldview, marginalizing them (Aikenhead, 1996).

From these excerpts it becomes clear that science education lacks space for discussing different epistemologies and ontologies, such as indigenous ways of knowing, and is also permeated with prejudices upon many social groups, representing an isolated and hostile environment for many students and for their cultures. Thus, the next stops in this tour, at middle elevations, respond to this challenge and include theoretical analyses that articulate oppression and practices that expand possibilities for *all* learners' participation in physics.

Grounding these analyses are immeasurable contributions to science by women, people of color and indigenous peoples that remain unacknowledged in History of Science, NOS, and PER. Astronomical, navigational, geographic and botanical knowledge and expertise in regions outside Europe both long predate Western colonizing forms of these sciences and contribute directly to Western findings (Semali *et al.*, 2015; Raj, 2018; and Ogunniyi, 2020). Stories of feminist and indigenous achievement and sexist, racist, and colonizing violence are routinely "silenced" in textbooks and science education (Ideland, 2018, p. 795). Through omitting these histories and sufferings, decontextualized physics education distorts the science and delegitimizes students.

Unlike the decontextualized "view from nowhere" idea, feminist standpoint theories emphasize that objectivity and knowledge are "situated" in the specifics of its contexts and participants. In fact, masculine culture underlies physics' "gender-neutral" veneer (Traxler *et al.*, 2016, p. 7). Subverting the dominance of masculinity, feminist education values: students who have been historically oppressed; empowers them; and acknowledges that physics is diminished without them (Rodriguez *et al.*, 2022). Challenging past educational research's implication of "why can't women be more like men?," PER feminists look to "change the culture" itself (Traxler *et al.*, 2016, p. 4). Feminist science learning and participation becomes ongoing, critical, complex, contextual (Barton, 1997) – requiring low elevation environments.

In harmony with feminist-situated contexts, in indigenous cultures "subject matter is properly examined and interpreted *contextually*" (Snively and Corsiglia, 2001, p. 11, emphasis added). While feminists emphasize bodily-situated experiences, indigenous experiences deeply situate in "place," respect land and resist colonization (Lowan-Trudeau, 2018). Analogous to multiple feminist standpoints, "indigenous ways of living in nature …require experiential processes" (Aikenhead and Ogawa, 2007, p. 553). These ways are relational, holistic, dynamic, place-based, systematically empirical, temporally cyclical, rational, revisited over time and validated through survival (Aikenhead and Ogawa, 2007). Observing life-spirit as inhering throughout nature, indigenous ways do not separate people from nature; instead, all beings practice mutual respect, reciprocity and responsibility (Snively and Corsiglia, 2001).

Through extended, communal experience with natural phenomena such as time and the cosmos, indigenous peoples develop understandings that are coherent and continue through story-telling and other contextual activities. These indigenous understandings bear commonalities and connections with Western science, while originating through—and continuing in—non-Western ways. Having appropriated indigenous understandings, benefited from the resulting technologies, and thereby identified insufficiency in Western methods, Western scientists and educators theorize approaches to collaborate with indigenous peoples. Rather than charting correspondences between indigenous and Western findings, ethnobiologists attend to such nuance as "partial overlap" between indigenous life-spirit in nature and recent Western consideration of plant cognition (Ludwig and El-Hani, 2020, p. 7). "Co-existence" among differing science practices requires opening awareness to the situated character of all knowledge and experience, and facing up to colonizing power imbalance (Snively and Corsiglia, 2001).

Feminist and indigenous practices open science participation in learners' lived experiences in relational dialogue with each other, history and nature. Through concurrent, respectful plurality in classroom participation, no single "view from nowhere" or oppressive stance prevails. Critical thinking brings differing contexts and timescales into relation. Critical acts for educators include confronting biases, opening conversations with indigenous students and elders, and facilitating mutual respect (Higgins, 2010). Integrating indigenous experience into physics education provides renewed views on science, allowing indigenous students to deepen their cultural values (Zidny *et al.*, 2020).

Through forming relationships of mutual respect, we cross the threshold making low elevation education possible. In summarizing diverse examples where students engage with feminism and indigenous experience, we illustrate the relational fluidity and learning at low elevations.

Low elevation education was transformative for one student of color, Melanie. A study observed Melanie longitudinally across her experiences in her urban US middle school science class. Struggling with science, Melanie was excluded by a dominant classmate. Classmates and teachers reacted by supporting her. Melanie transformed. Unlike before where she declined to participate in class, now Melanie coached her teammate saying "You're smart! So do your work now!" (Tan and Barton, 2008, p. 582). Encouraged by those around her, Melanie undertook new intellectual and social risks and developed personal agency.

Saying "Can we afford to wait?" for holistic science curriculum to become mainstream, researchers enacted their own vision by integrating feminist standpoint theory and constructivism into a college-level physical science course (Roychoudhury *et al.*, 1995, p. 899). Developing through students' lived experiences for learning, this course implemented extended projects and encouraged cooperation. Students initiated open-ended projects. Excitedly relating a lab on condensation with showering at home, one woman shared "I was explaining to my mom why the mirror gets fogged up!" (Roychoudhury *et al.*, 1995, p. 8). While applying classroom physics to day-to-day experiences, inversely, students realized how lived experience involves physics.

Students in the Gender PRO MINT program at Technische Universität Berlin apply feminist theory to analyze and critique how gender is practiced and institutionalized in STEM contexts where they are involved as students and researchers (Lucht and Mauß, 2015). Students then question their own biases in professional fields and apply their new perspective on gender in self-defined research projects. One project interviewed queer physicists and female physicists who migrated to Germany about their experiences of discrimination in physics (Lucht, 2021).

Theater, historical case studies, role playing, and student debate are among classroom methods that invite students to grapple with moral, feminist and ethical values (Allchin, 1999; and Chowdhury, 2018). Feminist texts are source material for impromptu dramatic improvisation in Feminist Theory Theater (FTT) pedagogy (Glutzman, 2017). Its "embodied, situated and distributed sense-making," allow participants to experience taking a stand (Aushana *et al.*, 2022, p. 18). While Western education conventionally separates morals from science (Allchin, 1999), feminist and Islamic morals figure in Malaysian and Indonesian science education (Tan, 1997; and Hindarto and Nugroho, 2018).

Crucially, feminist and social justice pedagogies mobilize students beyond the classroom (Brickhouse, 2001). Students and teachers apply science education to expose social injustices and undertake action (Yacoubian and Hansson, 2020). Speaking in Spanish, their community's language, not their school's, at a public event, students of color presented results of classroom labs analyzing local soil and power plants' detrimental impacts. Student investigator Odette reflected on how students identified injustice and became agents of change:

> The location of the power plants suggests environmental racism. … It was cool to have the science out for the people and I feel that it was even more important that it was…high school students who live in the community doing the science (Morales-Doyle, 2017, pp. 1050–1053).

Through respectfully and contextually addressing indigenous cultures, low elevation education widens students' growth through indigenous and academic domains (Abrams, 2014). Respectful practices acknowledge many pathways; honor "place;" involve community elders; encourage student voice, leadership and agency; and promote experiential learning (Antonellis, 2013).

Schools specifically for indigenous learners, where their culture pervades pedagogy and content, offer potential for such respectful coexistence practices, as illustrated by the next research examples. At a tribal college of the Tohono O'odham nation in Arizona, a white constructivist teacher developed and researched a culturally relevant physics course. Beforehand, by taking language classes with a tribal educator who used storytelling and constructivist methods, she aimed to build meaningful relationships with students and culture. Through dialogue welcoming cultural understandings, tribal students and teacher-researchers co-created knowledge around abstract physics concepts. One activity involved making a historical O'odham weapon, the atlatl. Connecting energy transfer, in physics, to techniques for releasing its dart, one student reflected:

> …you lose a whole lot of your energy … to go like this [move arm to spear-throwing position]. Then, where you just go like that [flick wrist to release dart from atlatl], … you are transferring more [energy] into the object (Antonellis, 2013, p. 224).

Extending this experience through initiating his own experiments, this student compared atlatl-launched darts with those released conventionally.

Elementary students at Niji Mahkwa School, Canada, an urban indigenous school equipped with modern science labs, engaged in their own culture while studying sound. They built and painted a ceremonial drum (Fig. 13.5). A musician elder visited, instructing students on the proper use of drums and leading them in ceremony.[1] Concurrently, students learned about sound waves, vibrations, and beats. Their indigenous understanding, that each drum is unique, partnered with a Fourier analysis demonstrating each drum's distinctive signal (Metz, 2017).

Pre-service physics teachers used indigenous games in the Philippines such as *shato* (gillidanda) and *luksong baka* (leapfrog) to learn about projectile motion and Newton's Second Law, respectively. Teachers' journaling documented how their thinking evolved while building physics understandings experientially (Morales, 2016).

African science education is pervasively colonized. Western science comes off as disjoint from, and meaningless to students' lives. Researchers' efforts for culture-affirming spaces face profound negativity toward indigenous culture among teachers. Documenting this and other obstacles, a Tanzanian pilot study proposed "Ufunifu," a community values framework for teacher education (Semali, 2013; and Semali *et al.*, 2015).

---

[1] The drum is a sacred instrument in indigenous culture. Thus, guidance on the building, use, and care of the drums by students was provided by an indigenous elder.

FIG. 13.5
Elementary students paint traditional indigenous images on a ceremonial drum at Niji Mahkwa School, Canada.

"Ubuntu," or socially-responsive African values, undergird South African post-apartheid curriculum mandates for culturally-responsive science—mandates that remain largely unfulfilled. Observing that indigenous youth lack words to render western science concepts, one project produces translations (Ogunniyi, 2020). Traditional African ways of mediating conflict partnered with science education argumentation methods in a teacher education course. These teachers then introduced indigenous knowledge through argumentation in their classrooms (Hewson and Ogunniyi, 2011). Another South African study analyzed three high school science teachers' implementations of indigenous education. Distinctive approaches emerged, confirming indigenous knowledge via science explanations, presenting indigenous knowledge side-by-side with Western science, and inviting students' sharing of indigenous practices. The last approach, coherent with this Tour's low elevation examples, captivated the students. In discussing the practical uses of animals, one student offered "pig's fat was smeared on lips for protection" (Naidoo and Vithal, 2014, p. 259). Responding, the teacher described pharmaceutical companies' profit from such knowledge. African herself, she went on to encourage culturally sustainable uses of animals.

Feminist and indigenous ways of knowing share values, experiences of oppression, and relational methods, provocative for simultaneous enactment in physics classrooms. Students develop voice, agency and understanding in low elevation experiences respectful of their identity, critical of oppression, and open to multiple methods. Feminist and indigenous experiences expand physics education through resiliency and values essential for addressing today's environmental, social and educational challenges.

13.10 CONCLUSION

Having traversed the landscape along three differing tours, we find concurrences, demonstrating greatest student participation and learning at low elevations. Our tours descend from decontextualized heights to moderate terrains, to fluid exchanges. The main character, the traveler-learner undergoing tour experiences, is the novice physics student. The student's view is intrinsically subjective: personal identity and life story guide every step; cultural heritage is the outfitting; grades, structural inequities and internalized traumas are weights burdening the student's backpack. At high elevations, students struggle to survive, undergoing training exercises. Where slopes are smoothed by cascades and pools, students find places to stand, rest, look around, discuss, take in the view and initiate excursions of their own.

By sea shores, marshes and mangroves, backpacks fall off. As their full selves emerge, students move freely, diving, exploring, dancing and reflecting. Their voices are spontaneous, laughing, questioning, wondering, singing, experimenting with new tones, in dialogue together and with travelers from all walks in the landscape, of times past and to come. Using their analyses of water's democratizing powers, students expand its flow routes, seed clouds for rain, sculpt fountains, unstop dams, and unleash intermingling and leveling waters. Collaboratively, students dissolve the struts of grading, structural inequities and trauma. They practice respect as the core of every relationship: among each other, teachers, historical, indigenous and future peoples, and nature. Respect for nature becomes shared as students do science by developing mutually balanced relationships with nature. Students learn by listening and reflecting, as well as by investigating, creating, experimenting, observing, and interacting. Students' voices, stories and reflections are the substance and spirit of educational works and are vibrant and visible in all research communications.

Descent from high elevations is not always possible under certain social frameworks and circumstances. However, we encourage readers, educators and students to review their local and larger landscapes, try new routes within it, introduce water features wherever you may be, and reshape the landscape while being open to being reshaped by it. Who are the main characters in your educational journeys? What actions will main characters take? What experiences and methods engage them, with nature's landscape, surrounding contexts, each other and people across all times, in mutuality and respect? While these questions are up to you to enact, seek out traveling companions, guides and scouts. The educational researchers whose work this review documents are alongside you to lend a hand, share a story, and invite continuing community.

ACKNOWLEDGMENTS

The editors of this volume encouraged our study. MIT's Edgerton Center supported it: Jim Bales, Sandi Lipnoski, and Helen Harlan. Hermann Schiffer Fernandes and Zoubeida Dagher responded thoughtfully to our draft. David Spicer joined our meetings. David Hammer's feedback is generative. Alvin Harvey shared his indigenous experience. We thank Alva Couch, Cindy Cheng, Peter Heering, Dietmar Höttecke, Xiao Huang, Yasmin Jayathirtha, Vassilis Koulountzos, Gopal Krishnamurthy, Michael Matthews, R. Ramanujam, Fanny Seroglou, Shailesh Shirali, Jinwoong Song, Zheran Wang, and Roland Wittje. In memory: Philip Morrison, Sukant Tripathy, Alanna Connors, and our dear departed ones.

REFERENCES

Abd-El-Khalick, F., *Handbook of Research on Science Education*, edited by L. Lederman and S. K. Abell (Routledge, 2014), Vol. II, pp. 635–664.

Abrams, E. *et al.*, *Handbook of Research on Science Education*, edited by L. Lederman and S. K. Abell (Routledge, 2014), Vol. II, pp. 685–710.

Adúriz-Bravo, A. and Pujalte, A. P., *Nature of Science for Social Justice*, edited by H. Yacoubian and L. Hansson (Springer, 2020), pp. 201–224.

Adúriz-Bravo, A., *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. Matthews (Springer, 2014), pp. 1443–1472.

Adúriz-Bravo, A., *Adapting Historical Knowledge Production to the Classroom*, edited by P. V. Kokkotas *et al.* (Sense Publishers, 2011), pp. 195–204.

Aikenhead, G. S. and Ogawa, M., Cult. Stud. Sci. Educ. **2**(3), 539–620 (2007).

Aikenhead, G. S., Stud. Sci. Educ. **27**(1), 1–52 (1996).

Allchin, D. (2010). Deating Galileo's Dialogue, 1633. SHiPS, see Retrieved February 28, 2022 from http://shipseducation.net/modules/phys/galileo-trial.htm.

Allchin, D., Sci. Educ. **95**(3), 518–542 (2011).

Allchin, D., Sci. Educ. **87**(3), 329–351 (2003).

Allchin, D., Sci. Educ. **21**, 1263–1281 (2012).

Allchin, D., Sci. Educ. **8**(1), 1–12 (1999).

American Institute of Physics (AIP) (2023). Teaching Guides and Educational Games on History of the Physical Sciences. Retrieved April 28, 2022 from https://www.aip.org/history-programs/physics-history/teaching-guides.

American Physical Society (APS) (2015). Statement on the Status of Women in Physics, see https://www.aps.org/policy/statements/15_2.cfm.

American Physical Society (2018). Diversity Statement. https://www.aps.org/policy/statements/08_2.cfm.

Antonellis, J. C., Unpublished doctoral dissertation (The University of Arizona, 2013).

Archer, L. *et al.*, Sci. Edu. **99**(2), 199–237 (2015).

Arons, A. B., Educ. Philos. Theory **20**(2), 13–23 (1988).

Arons, A. B., Am. J. Phys. **50**(1), 13–20 (1982).

Aushana, C. *et al.*, *STS Infrastructures, Platform for Experimental Collaborative Ethnography*, edited by S. Klein *et al.* (Society for Social Studies of Science, 2022) (last accessed April 4, 2022), see https://stsinfrastructures.org/content/feminist-theory-theater-workbook.

Avraamidou, L., J. Res. Sci. Teach. **57**(3), 311–341 (2020).

Barthelemy, R. S. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 010124 (2022).

Barton, A. C. and Tan, E., J. Learn. Sci. **19**(2), 187–229 (2010).

Barton, A. C., Curric. Inq. **27**(2), 141–163 (1997).

Batista, R. F. and Silva, C. C., Sci. Educ. **28**(9), 1135–1151 (2019).

Bell, R. *et al.*, Sci. Educ. **10**(1), 187–204 (2001).

Blum, S. D. and Kohn, A., *Ungrading: Why Rating Students Undermines Learning (and What to Do Instead)* (West Virginia University Press, 2020).

Brickhouse, N. W., J. Res. Sci. Teach. **38**(3), 282–295 (2001).

Bruner, J., *Actual Minds, Possible Worlds* (Harvard University Press, 1986).

Carchon, R. and Segers, D., *Historical Scientific Instruments in Contemporary Education*, edited by E. Cavicchi and P. Heering (Brill, 2021), pp. 139–157.

Cardinot, D. *et al.*, Sci. Edu. 1–33 (2022).

Carlone, H. B. and Johnson, A., J. Res. Sci. Teach. **44**(8), 1187–1218 (2007).

Carvalho, W. and Carvalho, C. A. B. (2002). Roleplays in Middle School Science Textbooks: A Significant Contribution to the History of Science Teaching. ERIC ED46963.

Casper, A. M. A. and Balgopal, M. M., Int. J. Sci. Educ. **42**(9), 1568–1584 (2020).

Cavicchi, E. M., Unpublished doctoral dissertation (Harvard University, 1999).

Cavicchi, E., Interchange **49**, 25–68 (2018).

Cavicchi, E., Interchange **42**(1), 21–50 (2011).

Cavicchi, E., *Historical Scientific Instruments in Contemporary Education*, edited by E. Cavicchi and P. Heering (Brill, 2021b), pp. 158–180.

Cavicchi, E., Interchange **45**(3), 185–204 (2014).

Cavicchi, E., *Looking and Listening for Learning*, edited by M. K. Delaney and S. J. Mayer (Teachers College Press, 2021a), pp. 129–145.

Chambers, D. W. and Gillespie, R., Osiris **15**, 221–240 (2000).

Chang, H., Sci. Educ. **20**(3), 317–341 (2011).

Cheney, I. and Shattuck, S. (2020). (Directors). *Picture a Scientist* (Film). Uprising Productions.

Chowdhury, M., J. Edu. Sci. **4**(2), 1–16 (2018).

Clough, M. P., Sci. Educ. **20**(7), 701–717 (2011).

Cobern, W. W. and Loving, C., *Nature of Science in Science Instruction*, edited by W. McComas (Springer, 2020), pp. 213–222.

Cobern, W. W., Int. J. Sci. Educ. **18**(3), 295–310 (1996).

Conant, J. B., *Harvard Case Histories in Experimental Science* (Harvard University Press, 1957).

Dagher, Z. R. and Erduran, S., Sci. Educ. **25**(1), 147–164 (2016).

Dagher, Z. R., *Nature of Science for Social Justice*, edited by H. Yacoubian and L. Hansson (Springer, 2020), pp. 41–58.

de Castro, R. S. and de Carvalho, A. M. P., Sci. Educ. **4**(1), 65–85 (1995).

Dewey, J., *Art as Experience* (Balch & Company, Minton, 1934).

Dewey, J., *Democracy and Education: an Introduction to the Philosophy of Education* (Free Press, 1916).

Dini, V. and Hammer, D., Phys. Rev. Phys. Educ. Res. **13**(1), 010124 (2017).

Duckworth, E., *The Having of Wonderful Ideas and Other Essays on Teaching and Learning* (Teachers College Press, 2006).

Dumit, J., *Picturing Personhood: Brain Scans and Biomedical Identity* (Princeton University Press, 2004), Vol. 2.

Dunlop, L. and de Schrijver, J., *Nature of Science in Science Instruction*, edited by H. Yacoubian and L. Hansson (Springer, 2020), pp. 223–237.

Erduran, S. and Dagher, Z. R., *Reconceptualizing the Nature of Science for Science Education: Why Does it Matter?*, edited by Z. Dagher and S. Erduran (Springer, 2014), pp. 1–18.

Feldman, J., *Grading for Equity: What it is, why it Matters, and how it can Transform Schools and Classrooms* (Corwin Press, 2018).

Freire, P., *Cultural Action for Freedom* (Harvard Educational Review, 1970).

Freire, P., *Pedagogy of the Oppressed*, edited by M. Ramos (Continuum, 2000).

Galili, I., Sci. Educ. **21**(9), 1283–1316 (2012).

Gandolfi, H. E., Cult. Stud. Sci. Educ. **14**(3), 557–567 (2019).

Gentzi, E. and Seroglou, F., *Re-introducing Science: Sculpting the Image of Science*, edited by F. Seroglou and V. Koulountzos (Grafima Publications, 2019), Vol. 38.

Gibbs, L., *Ungrading: Why Rating Students Undermines Learning (and What to do Instead)*, edited by S. D. Blum and A. Kohn (West Virginia University Press, 2020), p. 91.

Glashow, S. L., *From Alchemy to Quarks. The Study of Physics as a Liberal art* (Brooks/Cole, Pacific Grove, CA, 1994).

Glutzman, Y., *Imagined Theatres: Writings for a Theoretical Stage*, edited by D. Sack (Routledge, 2017), pp. 80–81.

Gorman, M. E. and Robinson, J. K., Sci. Educ. **7**(2), 173–201 (1998).

Hadzigeorgiou, Y. *et al.*, Sci. Educ. **21**(8), 1111–1138 (2012).

Hansson, L., *Nature of Science in Science Instruction*, edited by H. Yacoubian and L. Hansson (Springer, 2020), pp. 627–639.

Hansson, L. *et al.*, Phys. Educ. **54**(4), 045002 (2019b).

Hansson, L. *et al.*, Phys. Educ. **54**(5), 055008 (2019a).

Haraway, D., Fem. Stud. **14**(3), 575–599 (1988).

Hawkins, D., *The Informed Vision: Essays on Learning and Human Nature*, edited by D. Hawkins (Algora Publishing, 2002).

Heering, P. and Cavicchi, E., *Nature of Science in Science Instruction*, edited by W. McComas (Springer, 2020), pp. 609–626.

Heering, P. and Höttecke, D., *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. Matthews (Springer, 2014), pp. 1473–1502.

Heering, P. and Klassen, S., Phys. Educ. **45**(4), 382 (2010).

Heering, P. and Winchester, I., Interchange **46**(1), 1–3 (2015).

Heering, P., Sci. Educ. **9**(4), 363–373 (2000).

Heinicke, S. and Heering, P., Sci. Educ. **22**(3), 483–503 (2013).

Heinicke, S., Interchange **45**(3), 217–233 (2014).

Henke, A. and Höttecke, D., Sci. Educ. **24**(4), 349–385 (2015).

Hewson, M. G. and Ogunniyi, M. B., Cult. Stud. Sci. Educ. **6**(3), 679–692 (2011).

Higgins, M., Unpublished Masters thesis (Lakehood University, Ontario, 2010).

Hindarto, N. and Nugroho, S. E., J. Phys. Conf. Ser. **983**(1), 012002 (2018).

Höttecke, D., Sci. Educ. **21**(9), 1229–1232 (2012), see https://users.auth.gr/bkoul/hipst/hipst_docs.

Höttecke, D., Sci. Educ. **9**(4), 343–362 (2000).

Ideland, M., Sci. Educ. **27**(7), 783–803 (2018).

Jaber, L. Z. *et al.*, J. Res. Sci. Teach. **59**(2), 223–251 (2022).

Jardim, W. T. *et al.*, Sci. Edu. **30**, 609–638 (2021).

Johnston, J. S. *et al.*, *International Handbook of Research in History, Philosophy and Science teaching,* edited by M. Matthews (Springer, 2014) pp. 2409–2432.

Johnston, P. H. and Nicholls, J. G., Theory Pract. **34**(2), 94–100 (1995).

Keller, E. F., *Working It Out: 23 Women Writers, Artists, Scientists and Scholars Talk About Their Lives and Work*, edited by S. Ruddick and P. Daniels (Pantheon, 1977), pp. 78–91.

Khantine-Langlois, F. *et al.*, *Historical Scientific Instruments in Contemporary Education*, edited by E. Cavicchi and P. Heering (Brill, 2021), pp. 122–138.

Klassen, S. and Klassen, C. F., *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. Matthews (Springer, 2014), pp. 1503–1529.

Klassen, S., Interchange **37**(1–2), 31–62 (2006).
Klassen, S., Sci. Educ. **18**(3), 401–423 (2009).
Koliopoulos, D. *et al.*, Sci. Educ. Rev. **6**(2), 44–56 (2007).
Koulountzos, V. and Seroglou, F., *Adapting Historical Knowledge Production to the Classroom*, edited by P. V. Kokkotas *et al.* (Brill, 2011), pp. 213–227.
Koulountzos, V. *et al.*, *Re-introducing Science Sculpting the Image of Science*, edited by F. Seroglou and V. Koulountzos (Grafima Publications, 2019), p. 483.
Kubli, F., Sci. Educ. **14**(6), 501–534 (2005).
Larochelle, M. *et al.*, *Constructivism and Education* (Cambridge University Press, 1998).
Laux, K., Res. Sci. Technol. Educ. **36**(1), 111–129 (2018).
Lazos, P. *et al.*, *Historical Scientific Instruments in Contemporary Education*, edited by E. Cavicchi and P. Heering (Brill, 2021), pp. 105–121.
Lederman, N. G. and Lederman, J. S., *Handbook of Research on Science Education*, edited by N. G. Lederman and S. K. Avell (Routledge, 2014), Vol. II, pp. 614–634.
Lederman, N. G., J. Res. Sci. Teach. **29**(4), 331–359 (1992).
Lederman, N. G. and Lederman, J. S., Dis. Interdis. Sci. Edu. Res. **1**, 1–9 (2019).
Leone, M. and Rinaudo, M., Phys. Educ. **55**(3), 035013 (2020).
Lincoln, Y. S., Theory Pract. **34**(2), 88–93 (1995).
Lowan-Trudeau, G., *New Moral Natures in Tourism*, edited by S. R. Bryan *et al.* (Taylor & Francis, 2018), pp. 181–193.
Lucht, P. and Mauß, B., *Proceedings of the Annual Conference of the European Society for Engineering Education, SEFI Annual Conference* (SEFI, 2015), pp. 29–06.
Lucht, P., *The Human Rights-Based Approach to STEM Education*, edited by T. Tajmel *et al.* (Waxmann, 2021), pp. 147–168.
Ludwig, D. and El-Hani, C. N., J. Ethnobiol. **40**(1), 3–20 (2020).
Martin, B. E. and Brouwer, W., Sci. Educ. **75**(6), 707–722 (1991).
Mathai, S., *Exploring Hybrid Spaces Through an Informal Science Learning Programme* (Azim Premji University, 2017), Working Paper 8.
Matthews, M. R. *et al.*, *The Pendulum: Scientific, Historical, Philosophical and Educational Perspectives* (Springer Science & Business Media, 2005).
Matthews, M. R., *International Handbook of Research in History, Philosophy and Science Teaching* (Springer, 2014).
Matthews, M. R., *Science Teaching: The Contribution of History and Philosophy of Science* (Routledge, 1994).
Matthews, M. R., *Time for Science Education: How Teaching the History and Philosophy of Pendulum Motion Can Contribute to Science Literacy* (Plenum Press, 2000).
Maurines, L. and Beaufils, D., Sci. Educ. **22**(6), 1443–1465 (2013).
McComas, W. F., *Nature of Science in Science Instruction*, edited by W. McComas (Springer, 2020b), pp. 35–65.
McComas, W., *Nature of Science in Science Instruction. Science Philosophy, History and Education* (Springer, 2020a).
Melfi, T, (2016). (Director). Hidden Figures (Film). Levantine Films.
Metz, D., *Multiculturalidad y Diversidad en la Enseñanza de las Ciencias*, edited by M. Gatica (Bellaterra, 2017), pp. 117–126.
Metz, D. *et al.*, Sci. Educ. **16**, 313–334 (2007).
Milne, C., J. Res. Sci. Teach. **35**(2), 175–187 (1998).
Monk, M. and Osborne, J., Sci. Educ. **81**(4), 405–424 (1997).
Morales, M. P. E., Eurasia J. Math. Sci. Technol. Educ. **13**(5), 1377–1409 (2016).
Morales-Doyle, D., Sci. Educ. **101**(6), 1034–1060 (2017).
Morrison, P., *Nothing is Too Wonderful to be True* (American Institute of Physics Press, 1995).
Moshfeghyeganeh, S. and Hazari, Z., Phys. Rev. Phys. Educ. Res. **17**(1), 010114 (2021).
Moura, C. B. *et al.*, *Nature of Science for Social Justice*, edited by H. Yacoubian and L. Hansson (Springer, 2020), pp. 137–155.
Murphy, P. and Whitelegg, E., Curric. J. **17**(3), 281–305 (2006).
Naidoo, P. D. and Vithal, R., Afr. J. Res. Math. Sci. Technol. Educ. **18**(3), 253–263 (2014).
Neff, K. D. and Helwig, C. C., Cogn. Dev. **17**(3-4), 1429–1450 (2002).
Nissen, J. M. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(1), 010116 (2021).
Noddings, N. and Witherell, C., *Stories Lives Tell: Narrative and Dialogue in Education*, edited by C. Witherell and N. Noddings (Teachers College Press, 1991), pp. 279–280.
Ogunniyi, M. B., *Nature of Science for Social Justice*, edited by H. Yacoubian and L. Hansson (Springer, 2020), pp. 157–176.
O'Loughlin, M., Theory Pract. **34**(2), 107–116 (1995).
Ong, M., Soc. Probl. **52**(4), 593–617 (2005).
Osborne, J. F., Sci. Educ. **80**(1), 53–82 (1996).
Østergaard, E. *et al.*, Stud. Sci. Educ. **44**(2), 93–121 (2008).
Paparou, F., *Historical Scientific Instruments in Contemporary Education*, edited by E. Cavicchi and P. Heering (Brill, 2021), pp. 181–208.
Park, W. and Song, J., Sci. Educ. **27**(1), 39–61 (2018).
Park, W. *et al.*, Sci. Educ. **28**(9), 1055–1083 (2019).
Paul, C. A. and Webb, D. J., Phys. Rev. Phys. Educ. Res. **18**, 020103 (2022).
Philip, T. M. and Gupta, A., Rev. Res. Educ. **44**(1), 195–217 (2020).
Piaget, J. and Inhelder, B., *The Growth of Logical Thinking: From Childhood to Adolescence*, edited by A. Parsons and S. Milgram (Basic Books, 1958).
Piaget, J., *The Moral Judgment of the Child*, edited by M. Gabain (Free Press, 1965), see http://www.archive.org/details/moraljudgmentoft005613mbp.

Piaget, J., *The Origins of Intelligence in Children*, edited by M. Cook (International Universities Press, 1952).

Piliouras, P. *et al.*, Sci. Educ. **20**(7), 761–795 (2011).

Prescod-Weinstein, C., Signs: J. Women Cult. Soc. **45**(2), 421–447 (2020).

Prescod-Weinstein, C., *The Disordered Cosmos: A Journey Into Dark Matter, Spacetime, and Dreams Deferred* (Hachette, UK, 2021).

Project Physics Collection (2010). Internet Archive. Retrieved February 28, 2022 from https://archive.org/details/projectphysicscollection.

PSSC (2012–2022) Retrieved March 1, 2022 from https://archive.org/search.php?query=Physical%20Science%20Study%20Committee.

PSSC, *Physics* (D.C. Heath and Co, 1960).

Rader, K., Isis **111**(3), 568–575 (2020).

Radoff, J. *et al.*, Cogn. Instr. **37**(1), 73–92 (2019).

Raj, K., *Empires of Knowledge* (Routledge, 2018), pp. 269–293.

Ramadas, J., Indian Educ. Rev. **37**(2), 3–21 (2001).

Revel Chion, A. and Adúriz-Bravo, A., Sci. Educ. **31**(2), 269–291 (2022).

Roach, L. E. and Wandersee, J. H., School Sci. Math. **95**(7), 365–370 (1995).

Robertson, A. D. and Hairston, W. T., Phys. Rev. Phys. Educ. Res. **18**(1), 010119 (2022).

Rodriguez, M. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 013101 (2022).

Rogers, E. M., *Physics for the Inquiring Mind: The Methods, Nature, and Philosophy of Physical Science* (Princeton University Press, 1960).

Roychoudhury, A. *et al.*, J. Res. Sci. Teach. **32**(9), 897–924 (1995).

Rudolph, J., *Scientists in the Classroom: The Cold war Reconstruction of American Science Education* (Palgave, 2002).

Rutherford, F. *et al.*, *The Project Physics Course* (Rinehart and Winston, Holt, 1970).

Savage, C. *et al.*, Asia-Pac. J. Teach. Educ. **39**(3), 183–198 (2011).

Schvartzer, M. *et al.*, Sci. Educ. **30**(1), 165–179 (2021).

Seatter, C. S., Interchange **34**(1), 63–87 (2003).

Secules, S. *et al.*, J. Eng. Educ. **107**(2), 186–218 (2018).

Seiler, G., J. Res. Sci. Teach. **38**(9), 1000–1014 (2001).

Semali, L. M., J. Contemp. Issues Educ. **8**(2 (2013).

Semali, L. M. *et al.*, Cult. Stud. Sci. Educ. **10**(4), 865–889 (2015).

Shrigley, R. L. and Koballa Jr, T. R., Sch. Sci. Math. **89**(4), 293–298 (1989).

Silva, A. P. B. D. *et al.*, *Teaching Science with Context* (Springer, 2018), pp. 327–339.

Simmons, A. B. and Heckler, A. F., Phys. Rev. Phys. Educ. Res. **16**(2), 020125 (2020).

Skype A Scientist (2019) Retrieved February 28, 2022 from https://www.skypeascientist.com/.

Snively, G. and Corsiglia, J., Sci. Educ. **85**(1), 6–34 (2001).

Snively, G. J. and Williams, L. B., L1-Educ. Stud. Lang. Lit. 8, 109–133 (2008).

Sohr, E. R. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020157 (2020).

Song, J. and Joung, Y. J., *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. Matthews (Springer, 2014), pp. 2177–2215.

Spek, van der, T. M., *Historical Scientific Instruments in Contemporary Education*, edited by E. Cavicchi and P. Heering (Brill, 2021), pp. 243–262.

Sriprakash, A., Int. J. Educ. Dev. **30**(3), 297–304 (2010).

Stadermann, H. K. E. and Goedhart, M. J., Phys. Rev. Phys. Educ. Res. **17**(2), 020132 (2021).

Stefanidou, C. (2016, August). History and Philosophy of Science for Citizenship: The Case of "Life of Galileo" by B. Brecht. In *Abstracts 1st European Regional IHPST Conference, Flensburg*, pp. 22–25.

Stefanidou, C. *et al.*, Phys. Educ. **55**(3), 035027 (2020).

Stinner, A., Sci. Educ. **79**(5), 555–581 (1995).

Stinner, A., Interchange **37**(1), 19–30 (2006).

Stinner, A. *et al.*, Sci. Educ. **12**(7), 617–643 (2003).

Strommel, J., *Ungrading: Why Rating Students Undermines Learning (and What to do Instead)*, edited by S. D. Blum and A. Kohn (West Virginia University Press, 2020), pp. 25–41.

Tan, E. and Barton, A. C., Sci. Educ. **92**(4), 567–590 (2008).

Tan, S. K., Sci. Educ. **6**(6), 555–572 (1997).

Taylor, L. W. and Tucker, F. G., *Physics: The Pioneer Science* (Houghton Mifflin Co., 1941).

Teixeira, E. S. *et al.*, Sci. Educ. **21**(6), 771–796 (2012).

The Story Collider: Personal Stories About Science (2010). Retrieved February 28, 2022 from https://www.storycollider.org/.

Traxler, A. L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020114 (2010).

United Nations (2022). Indigenous Peoples at the United Nations, see https://www.un.org/development/desa/indigenouspeoples/about-us.html; https://www.un.org/esa/socdev/unpfii/documents/5session_factsheet1.pdf.

Vygotsky, L. S., *Thought and Language*, edited by E. Hanfmann and G. Vakar (Trans.) (MIT Press, 1962).

Wang, Z., *XXIX Scientific Instrument Symposium*, September 14 (Science Museum and Royal Museums Greenwich, 2020).

Webb, D. J. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020114 (2020).

Whitten, B. L., J. Women Minor. Sci. Eng. **18**, 115 (2012).

Wong, D. *et al.*, J. Res. Sci. Teach. **38**(3), 317–336 (2001).

Wong, S. L. *et al.*, *International Handbook of Research in History, Philosophy and Science Teaching*, edited by M. Matthews (Springer, 2014), pp. 2149–2175.

Woods-Townsend, K. *et al.*, Int. J. Sci. Educ. Part B **6**(1), 89–113 (2016).

Yacoubian, H. and Hansson, L., *Nature of Science for Social Justice* (Springer, 2020).

Yildirim, B. *et al.*, J. Balt. Sci. Educ. **20**(5), 740–758 (2021).

Zidny, R. *et al.*, Sci. Educ. **29**(1), 145–185 (2020).

SECTION
III PHYSICS TEXTBOOKS

Section Editors

Marika Kapanadze and Gabriela Jonas-Ahrend

CHAPTER

14

EXPECTATIONS ON PHYSICS TEXTBOOKS

Sascha Grusche, Alexander Strahl, and Katrin Bölsterli Bardy

Grusche, S., Strahl, A., and Bardy, K. B., "Expectations on physics textbooks," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 14-1–14-36.

14.1 INTRODUCTION

A textbook is defined as a pedagogically prepared book for students to work with individually to support their learning, and for teachers to support their teaching (Zwahr, 2006, p. 486). Some definitions refer only to printed books, whereas others also include electronic versions. Regardless, it is hard to identify consistent features across all textbooks (Bölsterli Bardy, 2014). Moreover, the demarcation between the term textbook and other terms - such as schoolbooks, teaching aids, teaching materials, educational materials, and tools for teaching - is vague (Bölsterli Bardy, 2015). For this chapter, we define the textbook as an analog or digital book containing detailed information about a subject for teachers and learners, optionally including supplementary materials and a teachers' guide.

In physics education, textbooks play a central role. They help teachers to teach, and students to learn (Merzyn, 1994). Some education researchers believe that textbooks strongly influence students' test performance (Valverde *et al.*, 2002). As such, textbooks must fulfill both teachers' and students' expectations (Bölsterli Bardy, 2015).

Teachers have various ways to use textbooks sporadically or systematically (Mikelskis, 2008, p. 57). This is because a physics textbook may serve multiple functions, namely, as the following:

- a book for learning,
- a book to work with,
- a resource for teaching material,
- a reference guide to physics,
- an experiment manual,
- a guide for physics projects, and
- a library of exercises

https://doi.org/10.1063/9780735425514_014

Experienced teachers tend to use a textbook selectively, while beginners tend to read page after page (Bölsterli Bardy, 2014). However, this cannot be generalized. Curriculum-makers consider physics textbooks as crucial tools for steering and reforming teaching (Mikelskis, 2008). Thus, producers recreate textbooks whenever a new curriculum is officially announced. If authorized textbooks for new curricula are lacking, teachers depend on their own idea of instruction, with the risk of not implementing the new curriculum (Bölsterli Bardy, 2014). Clearly, it is not only students and teachers who hold expectations about physics textbooks but also other people, ranging from curriculum-makers to textbook producers and education researchers (Bölsterli *et al.*, 2014).

However, expectations about physics textbooks have not been systematically gathered in the existing literature. Therefore, the purpose of this chapter is to provide a systematic literature review of physics textbook expectations research (PTER) worldwide.

In Sec. 14.2, we give a theoretical overview of textbook elements, expectation holders, educational levels, historical phases, and nation-specific features of textbooks. In Sec. 14.3, we formulate the research question. In Sec. 14.4, we describe our methods of searching for, selecting, and analyzing papers about PTER. In Sec. 14.5, we describe the results of our literature analysis by reviewing the introduction, methods, results and discussion sections of the analyzed articles. In Sec. 14.6, we discuss the contributions of PTER. In Sec. 14.7, we give a summary and outlook for PTER.

14.2 THEORY

Before we can analyze expectations about physics textbooks, we need to become aware of some important aspects: First, we need to understand the elements that compose a textbook. Second, we need to take into account who are the relevant expectation holders. Third, we need to distinguish between different educational levels. Fourth, we can identify different phases in the history of the textbook, each characterized by its own key expectation. Finally, we need to be aware that school curricula are nation-specific, and so are textbook expectations.

14.2.1 Structural elements of physics textbooks

A physics textbook is composed of the following structural elements: guiding elements, main text, highlighted text, figures, tables, experimental descriptions or instructions, and tasks (Merzyn, 1994).

Guiding elements include the table of contents and the index. The **table of contents** gives an overview of the content structure of the textbook. The **index** helps the reader to look up specific content within the book. The **main text** presents the whole learning material in depth. It covers observations, phenomena, laws, and theories of physics as well as related disciplines. Ideally, the main text outlines a thought process that leads to laws, conclusions and applications. **Highlighted text** is typically a statement of a physical law, an explanation of a concept, a risk warning, or a summary of the main text. A **figure** can be a photo, a realistic drawing, a schematic drawing, or a diagram. Photos and realistic drawings often refer to everyday life or classroom experiments. Schematic drawings help students in the scientific processes of abstraction and symbolic representation, especially in the context of experiments, complex devices, and mental models. Diagrams are used to represent functional relationships. A **table** lists data in a structured and easily accessible manner. **Experimental descriptions** refer to historical experiments that are hard to perform in the classroom or to demonstration experiments that the teacher might show. These descriptions are used within the textbook to close the gap between an introductory question and a given answer. In contrast, **experimental instructions** help the students do experiments themselves. **Tasks** are often at the end of a section or chapter and are important for students to consolidate what they have learned.

Overall, the content structure of the textbook should represent the course of a typical lesson, starting with motivation and problem statement, going through the phases of doing, analyzing, and discussing experiments, and finally wrapping up with application and exercise (Merzyn, 1994; and cf. Bölsterli Bardy, 2014).

14.2.2 Expectation holders

Physics textbooks are produced and used by different people, and all these stakeholders have different expectations about physics textbooks: Curriculum-makers, authors, publishers, teachers, students, and education researchers.

Curriculum-makers set the official framework for teaching and learning, including standards for textbook creation and evaluation. **Authors** write and illustrate textbooks. **Publishers** make textbooks available to the intended readership. **Teachers** and **students** use textbooks in class and at home to teach and learn. **Education researchers** may consider textbooks as an object of interest for their education research.

14.2.3 Education levels

When it comes to formulating expectations about textbooks, it is important to distinguish between different levels of education. Each nation has its own particular education system with different levels. However, most education systems can be subdivided into three major education levels: Primary level (e.g., elementary school), secondary level (e.g., high school), and tertiary level (e.g., university). The secondary level can be further divided into the lower and upper levels. Physics textbooks are expected to be tailored to one of those education levels (Khoja and Ventura, 1997)

At the **primary level**, books are primarily intended to introduce students to the subject of science and provide initial experience in the field. At the **lower secondary level**, physics textbooks are expected to include first formalizations and mathematizations, yet the level of visualization and reference to everyday life should still be very high. At the **upper secondary level**, abstraction and mathematization are expected to increase. At the **tertiary level**, very high mathematization and formalization are expected, and the number of pages in books increases enormously. From the primary to the tertiary level, the content and design of physics textbooks change greatly. The reading difficulty should always correspond to the respective levels of the learners (Bölsterli Bardy, 2014).

14.2.4 Historical phases

Each year, new physics textbooks are published. Evidently, the physics textbook is not written in stone, but is changing over time. From the first physics textbooks until today, we can outline some historical phases, based on incisive events in physics and physics education (see also, Calinger *et al.* 2019).

- The **Aristotelean era** aimed at finding the causes of natural change by pure observation and logical argumentation.
- The **Classical era**, inaugurated by Bacon, Comenius, Galilei and Newton, aimed at understanding nature by systematic experimentation and far-reaching mathematization.
- The **Humanistic era** aimed at unfolding the individual's potential by Humboldt's formal approach and Pestalozzi's material approach to education.
- The **Modern era** revolutionized our worldview regarding time, space, macrocosm and microcosm, based on scientists' development of relativity theory and quantum theory.
- The **Post-Sputnik era** (Strube, 1985; and Haugsbakk, 2013) aimed at improving conceptual understanding.
- The **PISA era** (Pons, 2012; and Haugsbakk, 2013) aimed at promoting competence and output orientation.
- The **Post-Covid-19 era** is aimed at extending digital learning.

The era in which the textbook is used has a significant impact on the *pedagogical aims* associated with the textbook. In turn, these aims influence the methods and content of teaching and learning, and therefore, they also have an impact on the textbook (as an example, Simon, 2013; and Holovko, 2016). Arguably, Galilei's *Discorsi e dimostrazioni matematiche* from 1638 is one of the first physics textbooks. It marks a turn from older textbooks that focus on reporting on physics research to modern textbooks that focus on supporting physics education.

14.2.5 Nation-specific features

Textbooks have different contents and structures depending on the national citizenship of the intended readership. Because each country has its own curriculum, physics textbooks must be nation-specific to represent that curriculum (Kahveci, 2010). Moreover, textbooks need to be nation-specific because each nation has its own official language, history, and culture.

14.3 RESEARCH QUESTION

With the aim of creating an overview of physics education research on textbook expectations, our main research question is

What expectations about textbooks have been identified in physics education research, and how has this research been performed?

Accordingly, the sub-questions are as follows:

a. What expectations are expressed in the introduction section of each paper, especially in the research gap and research question?
b. Which ways of investigating and analyzing expectations are described in the methods section of each paper?
c. What expectations are stated in the results section of each paper?
d. What expectations are expressed in the discussion section of each paper, especially in the recommendations for textbook creation, use, and research?

14.4 METHODS

To answer this research question, a systematic literature review was performed.

14.4.1 Methods of search and selection

14.4.1.1 Method of search

To find relevant literature on expectations about physics textbooks, keywords to search for and databases to search in were defined.

The English keywords to be used were defined on three levels of hierarchy. The first level defines the **subject** to be physics, the second level refers to **textbooks and similar terms**, and the third level relates to various **areas, forms and holders of expectations:**

1. physics
2. textbook(s), curriculum material(s), curricular material(s), teaching material(s)
3. author(s), better, best, characteristics, checklist(s), choice(s), choose, choosing, construction(s), content(s), create, creating, creation(s), creator(s), criteria, curricular, curriculum, demand(s), design(s), designing, develop, developing, development(s), edit, editing, editor(s), educator(s), envision, envisioning, equation(s), evaluation(s), exercise(s), expectancy, expectancies, expectation(s), experiment(s), figure(s), good, grid(s), guide(s), guideline(s), homework, idea(s),

ideal(s), illustration(s), index, indices, innovate, innovating, innovation(s), intention(s), judge, judging, key word(s), math, mathematical, motive(s), need(s), norm(s), normative, opinion(s), perspective(s), prefer, preference(s), publisher(s), quality, question(s), redesign(s), redesigning, re-design(s), re-designing, reform(s), reformation(s), reforming, requirement(s), review(s), standard(s), student(s), table(s), task(s), teacher(s), text(s), view(s), vision(s), want, writing.

For each search attempt, one keyword from each level was entered into the keyword search. The database to be used was **Google Scholar** because it is known for yielding a high number of search results for research publications. Some of the keyword searches were repeated with **the Web of Science**, leading to fewer search results. Therefore, Google Scholar was chosen as the database for all keyword searches. In Google Scholar, the search mode was chosen to be "allintitle:", meaning that all keywords must appear in the title of the research paper. The search was not limited to any particular time period, enabling articles from different eras to be found.

14.4.1.2 Method of selection

From all of the available search results in Google Scholar, relevant articles were selected step by step through a process of exclusion:

1. Excluding articles where the title does not contain all of the **keywords** specified for a given search attempt.
2. Excluding articles where the abstract is English, but the main text is **non-English**.
3. Excluding articles which are clearly **off-topic** according to the title.
4. Excluding articles which are clearly **off-topic** according to the abstract.
5. Excluding articles which are of **low academic quality of the research approach**, according to a quick read of the body of the paper.

After the selection process, 39 papers remained to be analyzed in detail, see Table 14.1.

14.4.2 Methods of analysis

14.4.2.1 Spreadsheet construction

To analyze the different sections of each paper, a spreadsheet with 43 columns was used.

The first three spreadsheet columns were used to list the **year of publication**, the **names of the authors**, and the **title of each paper**. The remaining columns of the spreadsheet were used to extract relevant information from the introduction, methods, results and discussion part of each paper.

Based on the introduction of each paper, the **research gap** and **research question** were quoted directly or indirectly.

Table 14.1
Chronological list of the papers to be analyzed.

No.	Author	Land	No.	Author	Land
1	Newton (1984)	England	21	Holovko (2016)	Ukraine
2	Schultz (1989)	USA	22	Ververs (2016)	NL
3	Strube (1989)	Australia	23	Artuso (2017)	Brazil
4	Barojas and Trigueros (1991)	Mexico	24	Bancong and Song (2018)	Indonesia
5	Duit *et al.* (1992)	Germany	25	Takaoğlu (2018)	Türkiye
6	Rodríguez and Niaz (2004a)	USA	26	Türk *et al.* (2018)	Türkiye
7	Rodríguez and Niaz (2004b)	USA	27	Aguiar and Garcia (2019)	Brazil
8	Marshall and Linder (2005)	ZA/Sweden	28	de Souza and Garcia (2019)	Brazil
9	Podolefsky and Finkelstein (2006)	USA	29	Handayani *et al.* (2019)	Indonesia
10	Rozina (2006)	Canada	30	Citra *et al.* (2020)	Indonesia
11	Ogan-Bekiroglu (2007)	Türkiye	31	Fitriah (2020)	Indonesia
12	Mahardika (2013)	Indonesia	32	Gumilar and Amalia (2020)	Indonesia
13	Trebien and Garcia (2013)	Brazil	33	Haryanto and Syam (2020)	Indonesia
14	Brajkovic (2014)	Spain	34	Lous and Garcia (2020)	Brazil
15	Fatoba (2014)	Nigeria	35	Mahardika *et al.* (2020)	Indonesia
16	Heiner *et al.* (2014)	Canada	36	Mufit *et al.* (2020)	Indonesia
17	Tesfaye and White (2014)	USA	37	Sipayung, (2020)	Indonesia
18	Slisko (2014)	Mexico	38	Wiyanto *et al.* (2020)	Indonesia
19	Klieger and Sherman (2015)	Israel	39	Zuza *et al.* (2020)	Spain
20	Martins and Garcia (2015)	Portugal/Brazil			

To further characterize the **research question**, each paper was analyzed according to the following aspects:

- **Expectation holders** (authors, curriculum-makers, publishers, researchers, students, teachers, or others)
- **Expectation areas** (contents, keywords/index, main text, highlighted text, figures, tables, experiments, tasks, the book as a whole, the teacher's manual, supplementary material, or others)
- **Educational level** (I = Primary, II = Secondary, III = Tertiary, I and II, II and III, I - III, or unspecified)
- **Subject** (physics, science, or other)
- Reference country

- **Reference continent** (Africa; Asia; Australia; Europe, including Russia; North America; South America; or multiple)
- **Reference era** (before Sputnik, after Sputnik, after PISA, after Covid-19, or other)

All papers were categorized as being either empirical, literature-based, or normative:

1. **Empirical** reports of expectations are based on some method of oral or written inquiry.
2. **Literature-based** reports of expectations are based on a methodical literature review.
3. **Normative** statements of expectations are based on no research method, neither on inquiry nor on a methodical literature review.

In case of empirical studies, the **method of investigation** was analyzed according to the following aspects:

- **The type of investigation** (expert group, interview, literature review, none in case of normative expectations by the authors of a paper, questionnaire, multiple, or other)
- **Details about the method of investigation** (e.g., origin, steps, purpose)
- **Number of textbooks**, if textbooks were used to investigate expectations
- **Sample type** (authors, curriculum-makers, publishers, researchers, students, teachers, others) if expectations were empirically investigated
- **Sample size**, if expectations were empirically investigated

The **method of analysis** of each paper was analyzed according to the type (qualitative, quantitative, or both) and details of analysis (e.g., origin, steps, purpose).

The **results section** of each paper was analyzed regarding the expectations expressed by various expectation holders in various expectation areas.

The **discussion section** of each paper was analyzed regarding recommendations for future actions by various actors (authors, curriculum-makers, publishers, researchers, students, teachers, others) in various action domains (textbook creation, textbook use, textbook research, or other).

14.4.2.2 Spreadsheet conventions

Depending on the focus of a given paper, statements about expectations may be found anywhere in the paper. The following conventions were used to fill the spreadsheet for analysis:

1. Expectations that are stated in the introduction section as a research gap or research question are quoted in the spreadsheet columns titled “gap” or “research question.”
2. Expectations stated elsewhere in the introduction section or in the theory section are quoted in the spreadsheet columns titled “results.”
3. Expectations from the results section are quoted in the spreadsheet columns titled “results.”
4. Expectations from the discussion section are quoted in the spreadsheet columns referring to recommendations for textbook creation, use, or research.

Table 14.2
Intraclass correlation (ICC) for the categorization of research questions and gaps.

	1 vs 2	1 vs 3	2 vs 3	ICC
Round 1	89.9%	76.0%	85.2%	0.582
Round 2	88.5%	96.2%	93.5%	0.618

14.4.2.3 Inductive categorization of the gathered information

To summarize the textbook expectations gathered in the spreadsheet, an inductive categorization was performed. The following procedure was used:

1. Quote or paraphrase the expectations stated in each paper separately.
2. Summarize each expectation.
3. Generalize similar expectations across multiple papers in a short phrase.
4. Express each generalized expectation in an inductively gained category with a particular name, abbreviation, definition, and anchor example.
5. Subsume different sets of categories under a few domains.

Steps 3 to 4 were done iteratively to reduce the overall number of categories while bringing them to a similar level of generalization. First, these steps of generalization and definition were done separately for the research gap/question, results, and discussion sections of all papers. Then, generalization and definition were done across all these paper sections.

The analysis of papers was divided among the three authors of this chapter. To check whether all three authors used the inductive categories consistently, some parts of each paper were analyzed by all three authors, namely, the research questions and research gaps. For this portion of the analysis, the intraclass correlation (ICC) was determined. After the first round, the 23 categories were revised (Table 14.2). ICC improved in the second round and can be classified as average (Koo and Li, 2016) or good (Cicchetti, 1994).

14.5 RESULTS

14.5.1 Countries of the PTER articles

PTER is limited to only a few countries, at least for papers written in English. Quite a number of PTER papers have been published in Indonesia, Brazil, the U.S., and Türkiye (Fig. 14.1). A cross-national study (Marshall and Linder, 2005; and Martins and Garcia, 2015) could only be found twice.

Regarding the reference era, none of the analyzed papers are about the **era before Sputnik**. Only 5 out of 39 papers (12.8%) are exclusively about the **era after Sputnik**. Out of the 39 analyzed papers, 31 (79.5%) are exclusively about the **era after PISA**. One paper (2.6%) is about **both** the era after Sputnik

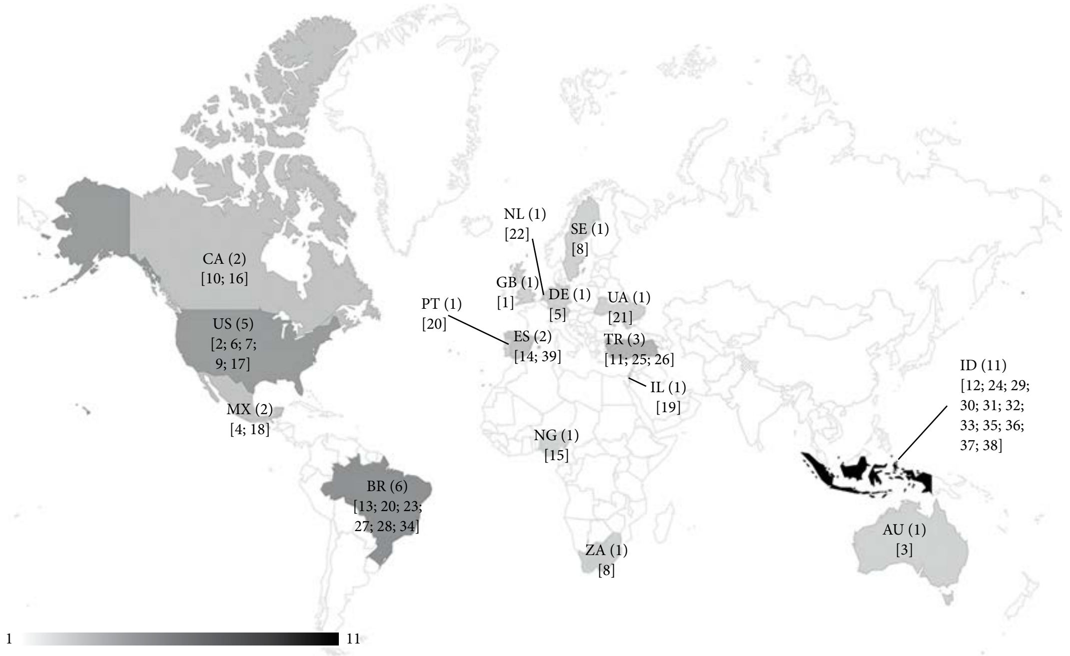


FIG. 14.1
Number of papers per country. The index number can be assigned to the year of publication, see Table 14.1. Meaning in the figure: Country name abbreviation (number of papers in the country) [number of the paper from Table 14.1].

and the era after PISA (Ververs, 2016). Only one paper (2.6%) is about the **era after Covid-19**. One paper (2.6%) gives a **historical overview** from 1893 until 2015.

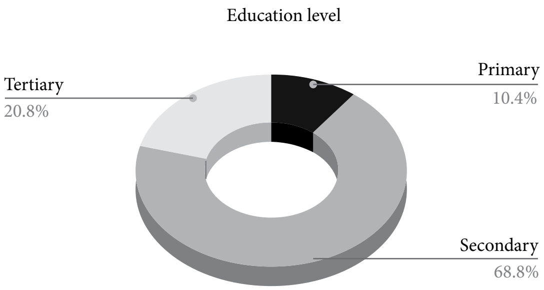


FIG. 14.2
The 39 papers that did PTER about the specified education levels.

Regarding the education level, most of the papers refer to **secondary education** (68.8%) (Fig. 14.2). **Primary** (10.4%) and **tertiary** education (20.8%) are rarely addressed. Note that some papers are about two educational levels, especially about secondary and tertiary education.

14.5.2 Inductively gained categories of expectations

In the following, the results of the inductive content analysis are shown. The inductively gained categories relate either to the content of a textbook, see Table 14.3, to the form of a textbook, see Table 14.4, or to an action

Table 14.3
Categories of expectations related to the content of a textbook.

Categories related to content	Abbreviation	Definition
Be course-matching	CM	The content and structure of the book should be [created or chosen (by the teacher) to be] aligned with the content and structure of the course.
Be nature-of-science-oriented	NO	The nature of science, including the contexts (historic, philosophical, social), methods and results of scientific insight, should be adequately represented.
Be competence-oriented	CO	The students' skills should be trained.
Be stereotype-free	SF	The textbook should be free of stereotypes regarding gender, race, ethnicity, occupation, etc.
Be curriculum-aligned	CA	The textbook should be created or chosen in line with the given educational curriculum.
Be interdisciplinary	ID	The textbook should incorporate contents from other disciplines besides physics, for example, art, biology, chemistry, religion …
Be student-oriented	SO	The textbook should address the students' perspective, including their needs, interests, prior knowledge, preconceptions, and cognitive level.
Be teacher-oriented	TO	The textbook should address the teacher's perspective, especially the teacher's needs and expectations regarding content, form and additional support.
Be context-based	CB	The physics content should be embedded in some kind of (real or fictional) context.
Be error-free	EF	The textbook should be free of errors in content or form.

Table 14.4
Categories of expectations related to the form of a textbook.

Categories related to form	Abbreviation	Definition
Be multi- representational	MR	A given content element should be represented in multiple structural elements, such as text, equations, and figures.
Be method-content aligned	MC	The content should be presented in a structured and methodical manner.
Be clearly laid-out	LO	The book should have a clear layout.
Fulfill external criteria	EC	Fulfill external criteria, such as low price and low weight.
Be fully digital	FD	The whole book should be available in digital format, in addition to or instead of analog.
Be research-based	RB	The content and form of the book should be based on insights from education research.

related to textbooks. Through the analysis, there are 10 content-related categories, 6 form-related categories, and 7 action-related categories were gained, yielding a total of 23 inductive categories of expectations. Note that the category *be research-based (RB)* can be content-related or form-related. Research-based means that it is based on empirical research on the topic.

14.5.3 Research gaps stated in the analyzed papers

In the following section, the unfulfilled textbook expectations found in the section "research gap" of the analyzed articles in PTER are summarized. The research gaps found are presented sorted thematically by their assigned category.

Most of the research gaps could be classified in the category of being *competence-oriented*. In general, textbooks are expected to be competence-oriented. However, PISA test results are bad (Türk *et al.*, 2018), the relationship between textbook use and student performance is unclear (Podolefsky and Finkelstein, 2006), students' creativity is rarely trained (Klieger and Sherman, 2015), students' foreign language skills are usually not trained in physics (Sipayung, 2020), new literacy and disaster literacy are hardly trained (Mufit *et al.*, 2020), textbooks rarely enable students to gain skills in problem solving and communication (Wiyanto *et al.*, 2020), and textbooks cannot replace the teacher (Newton, 1984).

In the category of being *student-oriented,* we found the following research gaps: students rarely read textbooks before class (Ververs, 2016), textbooks are expensive (Schultz, 1989), students' expectations about teaching are largely unknown (Marshall and Linder, 2005), cognitive domains are rarely systematically addressed (Barojas and Trigueros, 1991), it is hardly known what students find important in a textbook (Artuso, 2017), and the relationship between readability and student performance is hardly investigated (Fatoba, 2014). In the category of being *research-based,* quite a few research gaps have been found: the language difficulty and formality need to be defined (Strube, 1989), textbooks rarely fulfill all research-based expectations (Ogan-Bekiroglu, 2007), methodological requirements need to be developed (Holovko, 2016), textbooks are rarely based on insights into students' needs (Duit *et al.*, 1992), textbooks are rarely created through design-based research (Zuza *et al.*, 2020), and criteria for textbooks need to be applied (Lous and Garcia, 2020). Even though textbooks are expected to be aligned with the course, the following research gaps were stated in the category of being *course-matching*: it is difficult to find a fitting textbook because textbooks are changing over time (Tesfaye and White, 2014), there is a big variety of textbooks (Martins and Garcia, 2015), it is unclear how textbooks are introduced into class activities (Aguiar and Garcia, 2019), and criteria are needed for choosing a textbook (Lous and Garcia, 2020). There is a general consensus that textbooks are expected to be curriculum-aligned. However, there are some research gaps mentioned in the category of being *curriculum-aligned:* curricular goals and contents are sometimes unclear (Ververs, 2016), curricular guidelines do not always match classroom reality (Trebien and Garcia, 2013), textbooks do not always fit to a new curriculum (Takaoğlu, 2018), and the quality of textbooks does not always live up to curricular expectations (Haryanto and Syam, 2020). Some research gaps were mentioned within the category of being *multi-representational*: quantitative representations are rarely explained qualitatively

(Rozina, 2006), teacher candidates' multi-representational skills are rarely trained (Mahardika, 2013), textbooks often lack pictures and other representations (Mahardika *et al.*, 2020), and students' skills are rarely trained with multimodal representations (Handayani *et al.*, 2019; and Citra, Distrik *et al.*, 2020). Within the category of being *nature-of-science-oriented*, the following research gaps were mentioned: the history and philosophy of science are often neglected (Rodríguez and Niaz, 2004a), controversy as an aspect of science is often neglected (Rodríguez and Niaz, 2004b), and thought experiments are often neglected (Bancong and Song, 2018). In the category of being *context-based*, the analyzed papers mentioned the following research gaps: textbooks are full of artificial contexts (Slisko, 2014), textbook contents are rarely connected to students' reality (Lous and Garcia, 2020), and technologies used by students are hardly integrated with textbook use (de Souza and Garcia, 2019). In the category of being *stereotype-free*, it is mentioned that looking at possible stereotypes is of utmost importance, seeing that there are still stereotypes in textbooks concerning society, especially gender (Gumilar and Amalia, 2020). In the category of being *interdisciplinary*, it is mentioned that textbooks rarely include religion and local wisdom (Fitriah, 2020). In the category of being *fully digital*, it is stated that many textbooks are still offered in print, even though e-learning is becoming more and more popular (Brajkovic, 2014). In summary, the research gaps signify that there is a gap between the ideal and the real textbook.

14.5.4 Research questions posed in the analyzed papers

Analyzing the articles of PTER, we realized that expectations about textbooks are often the source of research questions. If textbooks should fulfill certain expectations, then certain research questions arise:

- If textbooks should be *student-oriented,* then: How do students use textbooks before class? (Heiner *et al.*, 2014). How, when, and why do students use textbooks? (Newton, 1984; and Podolefsky and Finkelstein, 2006). What are students' expectations about teaching? (Marshall and Linder, 2005). How are diverse cognitive domains addressed in textbooks? (Barojas and Trigueros, 1991). What are the most important features, topics, and foci of textbooks in the students' opinion? (Artuso, 2017). Do rural and urban students see different readability levels? Do male and female students perform differently with a given textbook? Do different textbooks lead to different performance? (Fatoba, 2014). Should local wisdom and religion be integrated into physics textbooks? (Fitriah, 2020).
- If textbooks should be *competence-orientated,* then: On which skill level are tasks regarding Bloom's taxonomy and PISA? (Türk *et al.*, 2018). Which types of tasks are included? (Takaoğlu, 2018). How is creativity promoted? (Klieger and Sherman, 2015). How is hypothetical-deductive reasoning promoted? (Mahardika *et al.*, 2020). Is disaster literacy trained? (Mufit *et al.*, 2020). Can communication skills be trained with physics textbooks? (Wiyanto *et al.*, 2020). How competence-based are the textbooks? (Ogan-Bekiroglu, 2007).
- If textbooks should be *curriculum-aligned*, then: How do curricular guidelines influence textbooks? (Ververs, 2016). Which evaluation criteria should be used to find textbooks in line with the course and the curriculum? (Trebien and Garcia, 2013). Do curriculum-based textbooks conform to international standards, for example, with regard to laboratory activities? How is the quality of

curriculum-based textbooks? (Haryanto and Syam, 2020). Is the textbook content in line with the curriculum? (Lous and Garcia, 2020).

- If textbooks should be *multi-representational*, then: Why and how to balance qualitative and quantitative representations? (Rozina, 2006). How can textbooks support teacher candidates' multi-representational skills? (Mahardika, 2013). How can students' reasoning be improved through multiple representations? (Mahardika *et al.*, 2020). Can skills be trained with multimodal representations? (Handayani *et al.*, 2019; and Citra *et al.*, 2020). Are the simulations suggested in textbooks appropriate? (de Souza and Garcia, 2019).
- If textbooks should be *research-based*, then: What are criteria for assessing textbook style? (Strube, 1989). Which quality characteristics can be derived from a historical analysis of textbooks? (Holovko, 2016). How can textbooks be student-oriented and still be competitive in the market? (Duit *et al.*, 1992). Does design-based textbook creation improve students' performance? (Zuza *et al.*, 2020).
- If textbooks should be *course-matching*, then: How to choose textbooks? (Schultz, 1989). Which textbooks do teachers use, and how do they rate them? (Tesfaye and White, 2014). How do textbooks influence the classroom curriculum? (Aguiar and Garcia, 2019).
- If textbooks should be *nature-of-science-oriented,* then: How are scientific models represented? (Rodríguez and Niaz, 2004a). How is scientific controversy represented? (Rodríguez and Niaz, 2004b). Are thought experiments included? (Bancong and Song, 2018).
- If textbooks should be *stereotype-free,* then: How are women and men portrayed? (Gumilar and Amalia, 2020).
- If textbooks should be *teacher-oriented,* then: Which factors make teachers choose a specific textbook? (Martins and Garcia, 2015).
- If textbooks should be *context-based,* then: What are causes and effects of artificial tasks? (Slisko, 2014).
- If textbooks should be *error-free,* then: How does the method of translation affect the accuracy of textbooks? (Sipayung, 2020).
- If textbooks should be *fully digital,* then: How are interactive e-books created? (Brajkovic, 2014).

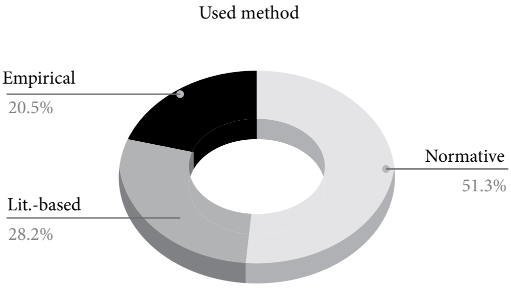


FIG. 14.3
The 39 papers used the specified methods to perform their PTER.

14.5.5 Methods described in the analyzed papers

14.5.5.1 Methods for investigating expectations

In the analyzed papers, expectations have often been stated **normatively**, without any basis in any method of investigation. This was the case for 20 out of 39 papers (51.3%), see Fig. 14.3.

The preferred method for investigating expectations has been a **literature review**. This has been done in 11 out of 39 papers (28.2%). Only rarely have expectations been investigated first-hand using some methods of **inquiry**. This has been done for 8 out of 39 papers (20.5%), always in the form of a questionnaire, except for one where an expert group was used. In addition to one of the questionnaires, an interview has been used (Martins and Garcia, 2015). In another case, the questionnaire was supplemented by electronic logs (Heiner *et al.*, 2014).

14.5.5.2 Methods for analyzing expectations

In the analyzed PTER papers, researchers have used methods of **qualitative content analysis** to analyze expectations, sometimes supplemented by a quantitative analysis for given categories.

14.5.6 Results reported in the analyzed papers

In this chapter, we present the expectations uttered in the analyzed articles of PTER in Sec. 14.5. The expectations are sorted by the expectation holders and by the structural elements of physics textbooks suggested by Merzyn (1994), see Chapter 2.1.

14.5.6.1 Expectations by curriculum-makers

Curriculum makers have expressed expectations regarding the contents, the main text, experiments, and supplementary materials.

Curriculum-makers expect contents to be *competence-oriented* by promoting students' skills in thinking, inquiry, design, modeling, and judging (Ververs, 2016). Moreover, contents should be *nature-of-science-oriented* by showing the relationships between physics, technology, and society, by discussing knowledge development in physics, and by developing and using scientific ideas. Contents should be *student-oriented* by providing a sense of purpose, by taking account of students' ideas, and by engaging students with relevant phenomena (Ververs, 2016). Finally, contents should be *context-based*, presenting physics concepts in context (Trebien and Garcia, 2013; and Ververs, 2016).

Curriculum-makers expect the main text to be *nature-of-science-oriented* by including derivations of important formulas (Trebien and Garcia, 2013).

Curriculum-makers expect experiments to be *course-matching*, being doable in the classroom, and *competence-oriented*, being accompanied by safety warnings (Trebien and Garcia, 2013).

Curriculum-makers expect supplementary materials to be *nature-of-science-oriented* by including derivations of important physics equations (Trebien and Garcia, 2013).

14.5.6.2 Expectations by authors/publishers

In none of the analyzed articles of PTER, expectations by authors or publishers have been found.

14.5.6.3 Expectations by teachers

Teachers have stated expectations about contents, keywords, the main text, highlighted text, figures, experiments, tasks, the book as a whole, and supplementary materials.

Teachers expect contents to be *student-oriented*: Contents should be meaningful, address the students' interests, and explain at the students' cognitive level. The text should be written at the students' reading level, providing technical language and pronunciations of foreign words in parentheses (Ogan-Bekiroglu, 2007). Content should also be *competence-oriented*, especially promoting critical thinking. They should be *curriculum-aligned*, being consistent with the scope and sequence of the curriculum. The textbook should be *multi-representational*, with many forms of representation, especially visualization. It should be *error-free*, especially regarding orthography. Content should be *nature-of-science-oriented*, by characterizing science as inquiry, by showing how science, technology and society interact, by treating the history of science, by highlighting physical laws (Ogan-Bekiroglu, 2007), and by showing that physics is more than math (de Souza and Garcia, 2019). The textbook should be *course-matching*, having the same topics as the course, and ideally no other topics (Schultz, 1989; and Martins and Garcia, 2015).

Teachers expect keywords to be *clearly laid-out*, with the name of each textbook unit being written on each page (Ogan-Bekiroglu, 2007).

Teachers expect the main text to be *student-oriented* by going from easy to difficult (Schultz, 1989), by having a suitable degree of complexity (Tesfaye and White, 2014), and by demonstrating how formulas are derived. Furthermore, the text should be *clearly laid-out*, with formulas being emphasized (Ogan-Bekiroglu, 2007). The main text should be *context-based* by starting with everyday examples, and *competence-oriented* by having clear and mathematically direct examples (Schultz, 1989).

Teachers expect highlighted text to be *clearly laid-out*, for example, by making important terms bold or italic, and by writing interesting headlines (Ogan-Bekiroglu, 2007). Further, the highlighted text should be *competence-oriented*, indicating the core of the course (Schultz, 1989).

Teachers expect figures to be *clearly laid-out* by being clear and colorful, strategically placed on the same page as the relevant text, featuring a millimeter scale in the case of a graph, and represented in a list of figures. Importantly, figures should be *student-oriented* by sparking interest, being understandable (with necessary quantities and units being given), and helping students process information. Figures should be *context-based*, with photos being up-to-date, and *error-free*, with proper captions. Figures should be *research-based*, relevant to the topic, and *content-method-aligned*, with illustrations about related rules being shown together. Overall, figures should be part of a *multi-representational* whole, for example, with photographs showing materials for experiments (Ogan-Bekiroglu, 2007).

Teachers expect experiments to be *nature-of-science-oriented*, with tasks directing students stepwise from hypothesizing to collecting data, and questions triggering critical thinking, observation, and investigation. Regarding experiments, textbooks should be *content-method-aligned* by asking students to discuss and draw conclusions. The experiments should be *student-oriented*, with some that can be done at home. Experiments should be *course-matching*, with safety rules being given. Finally, experiment-related texts should be part of a *multi-representational* whole, with illustrations supporting the explanation (Ogan-Bekiroglu, 2007).

Teachers expect tasks to be *competence-oriented* by promoting creative processes (Klieger and Sherman, 2015), by ranging from one-step to multi-step tasks, by having answers included, by avoiding unit conversions (Schultz, 1989), by supporting review, by including lab exercises, and by making students think critically and investigate. Moreover, tasks should be *student-oriented*, by going from easy to difficult, and by diagnosing students' alternative conceptions (Ogan-Bekiroglu, 2007).

Teachers expect the book to be *content-method-aligned* by treating the topics systematically, by supporting explanations with teaching strategies, by establishing a clear relationship between chapters and topics, by featuring marginal glosses (for questions, definitions, and main ideas), by being internally consistent, by having structured explanations that prevent alternative conceptions, and by having a clear organization (Ogan-Bekiroglu, 2007).

The textbook should be *context-based* by being up-to-date, by including projects, by addressing real-world situations, by explaining machines or tools, by showing physics-related jobs, and by relating to everyday life. The book should be *student-oriented* by considering students' interests, by addressing their abilities and needs, by being useful, by indicating pronunciations of foreign words, and by prioritizing students' demands over teachers'. At the same time, the book should fulfill *external criteria* such as being lightweight, being printed well on good paper, having pages that are small but not overfilled, and being durable. The whole book should be *clearly laid-out*, looking aesthetic, having good contrast between paper and print, with easy-to-read font size and line spacing, with margins for note-taking, and with an attractive cover. The book should be *curriculum-aligned*, consistent with the curriculum, or oriented toward the curriculum. It should be *stereotype-free*, especially regarding gender. To be *nature-of-science-oriented*, the textbooks should include experiments and research assignments. The whole book should be *error-free*; in particular, it should be scientifically accurate. Overall, the book should be *competence-oriented*, developing the students' cognitive, affective, and psychomotor skills. The whole book should be *research-based* in terms of its methodological treatment and written by competent authors with a related bachelor's degree (Ogan-Bekiroglu, 2007).

Teachers expect supplementary materials to be available, ideally as part of the textbook. They should be *course-matching*, for example, by enabling a "flipped classroom." Supplementary materials should be *competence-oriented*, addressing specific challenges. Ideally, they should be *fully digital* and available online. Moreover, additional resources should be used, such as teacher-made videos, magazines, the internet, and computer programs (Tesfaye and White, 2014).

14.5.6.4 Expectations by students

Students have stated expectations about contents, the main text, highlighted text, figures, experiments, tasks, the book as a whole, and supplementary materials.

Students expect contents to be *student-oriented* in many ways. The textbook should give many examples for concepts, provide advanced content, cite links (Artuso, 2017), be suitable for the students' social, emotional, and cognitive development level, address students' available skills, motivate for learning, ask for active involvement, build on the students' understanding of the material and encourage critical thinking (Fitriah, 2020). Importantly, the contents should be *context-based* by referring to media, society, art, comics, daily situations, the human body, nature, technology (Artuso, 2017), and real life (Fitriah, 2020). In addition, the book should be *nature-of-science-oriented* by promoting critical thinking, by discussing social, philosophical, and historical aspects of physics (Marshall and Linder, 2005), by including biographies, by outlining the historical context, by focusing on math, and by tracing the progress of scientific concepts. At the same time, the contents should be *interdisciplinary* by incorporating aspects from sports (Artuso, 2017), religion, economy, and other knowledge domains (Fitriah, 2020). Exercises should be *competence-oriented* beyond physics skills by training personal and social skills. Moreover, the book should be *curriculum-aligned*, for example, by being free of pornography (Fitriah, 2020). The contents should be *course-matching* and enable students to prepare for class (Heiner *et al.*, 2014). To be fully *multi-representational,* the book should also include simulations (de Souza and Garcia, 2019).

Students expect the main text to be *course-matching* because they mainly use the textbook after class: to answer questions and do exercises, to review, and as supplementary reading (Newton, 1984). To be *content-method-aligned*, the text should be coherent in its presentation, consistent in the use of terminology, and consistent in the use of symbols. In addition, it should be *error-free* regarding grammar and spelling (Fitriah, 2020). The text should be *research-based* to be short yet profound; and *student-oriented* to be enjoyable to read (Artuso, 2017).

Students expect the highlighted text to be *content-method-aligned*, for example, with supporting text being in boxes (Artuso, 2017).

Students expect figures to be part of a *multi-representational* whole by matching the text (Fitriah, 2020), being rich in synthetic and visual information, and appearing throughout the textbook. Furthermore, Figures should be *course-matching*, for example by providing a diagrammatic overview (Artuso, 2017); and *error-free*, for example by being numbered correctly (Fitriah, 2020).

Students expect experiments to be *student-oriented* by being easy to do (Artuso, 2017).

Students expect tasks to be *competence-oriented* by including examples for a problem as well as many types of exercises (Fitriah, 2020), including conceptual, numerical, and admission test exercises. Moreover, tasks should be *course-matching*, for example as part of a group discussion (Artuso, 2017).

Students expect the book as a whole to be *content-method-aligned* by having a presentation that is consistent and coherent across sections, subsections, and paragraphs. Overall, the book should be *clearly laid-out*, especially in the sequence of presentation, and *multi-representational*, having some variety of presentation (Fitriah, 2020). To be *student-oriented*, the book should include summaries and diagrams (Artuso, 2017), and only contain relevant elements (Fitriah, 2020). It should be *competence-oriented*, enabling students to solve homework problems and study for exams (Podolefsky and Finkelstein, 2006). The book should be *error-free*, especially free of conceptual errors, and it should fulfill *external criteria* such as being lightweight (Artuso, 2017).

Students expect supplementary materials to be preferably *fully digital*, in the form of simulations (de Souza and Garcia, 2019), digital content on the internet, and other multimedia content (Artuso, 2017). Supplementary materials should be *course-matching,* similar to the main text (Newton, 1984).

14.5.6.5 Expectations by physics education researchers

Researchers have stated expectations about contents, key words, the main text, figures, experiments, tasks, the book as a whole, and supplementary materials.

Researchers expect contents to be *competence-oriented*, by aiming at skills for everyday life (Duit *et al.*, 1992), by fostering communication and collaboration skills (Wiyanto *et al.*, 2020), by supporting cognitive and creative activities along individual learning paths (Holovko, 2016), by exercising language skills, cultural and political values (Aguiar and Garcia, 2019), by offering thought experiments (Bancong and Song, 2018), by promoting desirable behavior, problem-solving skills, and content knowledge (Fatoba, 2014), and by training skills in thinking, inquiring, designing, modeling, and judging (Ververs, 2016).

Textbooks should be *student-oriented*, establishing an imaginary dialogue between the author and the student (Holovko, 2016), starting from students' everyday experiences, going from everyday conceptions to physics conceptions, explicitly discussing students' preinstructional conceptions (Duit *et al.*, 1992), and showing the relationship between scientific and everyday knowledge (Lous and Garcia, 2020). Moreover, textbooks should catch the students' interest (e.g., with simulations) (de Souza and Garcia, 2019), explore new areas (Barojas and Trigueros, 1991), be sequenced from easy to difficult, address students' interests, enable conceptual assimilation, provoke conceptual accommodation, address diverse cognitive domains according to Piaget (Barojas and Trigueros, 1991), and provide different levels of exercises (Fatoba, 2014). Textbooks should use a clear language, motivate (Fatoba, 2014), provide a sense of purpose, consider students' ideas and engage students with relevant phenomena (Ververs, 2016).

At the same time, textbooks should be *nature-of-science-oriented*, by integrating the history and philosophy of science (Lous and Garcia, 2020), by discussing epistemological models (Aguiar and Garcia, 2019), by representing the scientific canon (Fatoba, 2014), by treating fundamental laws of physics (Ververs, 2016), by showing the relations between physics, technology, and society (Trebien and Garcia, 2013; and Ververs, 2016), by explaining knowledge development in physics, and by developing and using scientific ideas (Ververs, 2016). Textbooks should be *context-based* by starting from students' everyday experiences (Duit *et al.*, 1992), by contextualizing scientific epistemology (Lous and Garcia, 2020), by integrating local wisdom (Fitriah, 2020), by including disaster literacy (Mufit *et al.*, 2020), and by presenting concepts in contexts (Ververs, 2016). Moreover, authors should make textbooks *research-based*, by considering the model for concept acquisition in science, by combining educational aspects such as the nature of science, gender, and constructivism (Duit *et al.*, 1992), by using pedagogical models (Aguiar and Garcia, 2019), and by questioning the traditional textbook structure from an educational and scientific perspective (Strube, 1989). Content-wise, physics textbooks should be *interdisciplinary*, training friendly behavior toward humans and the environment (Duit *et al.*, 1992), integrating religion (Fitriah, 2020), and incorporating aspects from other disciplines (Trebien and Garcia, 2013). To be *content-method-aligned*, the book should provide a complete methodological system, set clear learning goals, be structured according to educational principles (Holovko, 2016), and outline learning paths (Barojas and Trigueros, 1991). Physics textbooks should be *curriculum-aligned* by embedding the curriculum (Lous and Garcia, 2020), and by fulfilling an ideological function (Aguiar and Garcia, 2019). They should be *stereotype-free*, yet address different genders and not represent males and females as being equal (Gumilar and Amalia, 2020). To be *multi-representational*, textbooks should contain equations (Lous and Garcia, 2020) and various other forms of representation (Mahardika, 2013; andMufit *et al.*, 2020).

Researchers expect keywords to be *nature-of-science-oriented* by being precisely defined within a scientific context and in relation to other scientific terms. Moreover, keywords should be part of a *multi-representational* whole, being defined through equations and words, possibly with metaphors (Strube, 1989).

Researchers expect the main text to be *nature-of-science-oriented* by including historical texts (Lous and Garcia, 2020), by discussing the scope and limits of scientific models, by explaining how scientific knowledge is generated and disseminated (Trebien and Garcia, 2013), and by emphasizing that experimental insights are preliminary, and that inductive and deductive methods interact (Rodríguez and Niaz, 2004a; and Rodríguez and Niaz, 2004b). The main text should be *research-based*, balancing between a traditional and pedagogical order of presentation. At the same time, it should be *context-based*, balancing between scientists and students. Moreover, it should be *content-method-aligned*, balancing between cold and warm writing styles, between scientific precision and conceptual development, and between strict and flexible ways of presenting information (Strube, 1989). The text should be *competence-oriented*, including preparation questions and sample problems (Türk *et al.*, 2018). It should be *error-free*, for instance in case of textbook translation (Sipayung, 2020). To make the text *student-oriented*, authors should use the approachable writing style of popular science books (Duit *et al.*, 1992).

Researchers expect figures to be *context-based*, depicting daily life and experience, and *nature-of-science-oriented*, illustrating the history of physics (Lous and Garcia, 2020). Figures should be part of a *multi-representational* whole by being one of many forms of representation (Handayani *et al.*, 2019) and by being consistent with the text. Figures should be *content-method-aligned*, having a clear purpose; they should be *clearly laid-out*; and they should be *error-free*, with precision (Trebien and Garcia, 2013).

Researchers expect experiments to be *course-matching*, doable in the classroom and posing no safety risks (Trebien and Garcia, 2013). To be *student-oriented*, the textbook should contain hands-on experiments based on everyday tools, and experiments that provoke a cognitive conflict. Experimental descriptions should be *content-method-aligned*, with an interplay between theory and experiment, and *context-based*, oriented toward phenomena from the students' lives. Experiments should be *competence-oriented*, promoting diverse skills (Schultz, 1989). To be *nature-of-science-oriented*, experiments should be presented in their historical context and theoretical framework, and with alternative interpretations of data (Rodríguez and Niaz, 2004a; and Rodríguez and Niaz, 2004b).

Researchers expect tasks to be *competence-oriented,* promoting problem-solving skills (Trebien and Garcia, 2013), inspiring an investigative attitude (Trebien and Garcia, 2013), and enabling students to enact known methods (Holovko, 2016). Tasks should be *context-based*, related to students' lives (Slisko, 2014), and authentic to make sense to students (Slisko, 2014). Ideally, tasks should be *nature-of-science-oriented*, including tasks in which students become researchers of others' concepts (Duit *et al.*, 1992). To be *multi-representational* and *content-method-aligned*, various types of tasks should appear in various parts of the textbook (Takaoğlu, 2018).

Researchers expect the book as a whole to be *course-matching,* balancing between curricular alignment and textbook diversity (Lous and Garcia, 2020), and being chosen in line with the content and structure of the course (Tesfaye and White, 2014). The book should be *research-based*, with design-based research promoting conceptual understanding, and determining the type of student activities (Zuza *et al.*, 2020).

To be *content-method-aligned*, the book should have methodological consistency (Lous and Garcia, 2020) and visual elements (such as fonts, headings, and figures) that support the content (Trebien and Garcia, 2013). The book should fulfill *external criteria* such as being available (Barojas and Trigueros, 1991), and being lightweight (Brajkovic, 2014). Moreover, the textbook should be *multi-representational* (Citra *et al.*, 2020), balancing between qualitative and quantitative representations (Rozina, 2006). It should be *error-free* (Lous and Garcia, 2020), and *nature-of-science-oriented* with a formal style (Strube, 1989). The book should be *student-oriented*, serving as the students' guide through the world of physics (Duit *et al.*, 1992). Ideally, there is a version of the book that is *fully digital*, with students being able to choose between printed and electronic versions of the book (Brajkovic, 2014).

Researchers expect supplementary materials to be *multi-representational*, with the textbook pointing at educational websites (de Souza and Garcia, 2019), and with an e-book containing videos, animation, sound, and web content. Ideally, the supplementary materials are *fully digital*, with multimedia being integrated in an e-book (Brajkovic, 2014).

14.5.7 Recommendations given in the analyzed papers

In the discussion part of the analyzed articles of PTER, several recommendations have been given by the researchers for textbook creation, textbook use, and textbook research.

14.5.7.1 Recommendations for textbook creation

Regarding content, researchers recommend textbook creation to be *nature-of-science-oriented*. For that, authors are advised to integrate different modes of physics (Tesfaye and White, 2014); to include scientific controversy and methodology (Rodríguez and Niaz, 2004b); to refer to the history and philosophy of science (Rozina, 2006), to describe important ideas and discoveries and include historical approaches (Ogan-Bekiroglu, 2007), and even to insert errors in scientific thinking. Overall, authors should try to give a correct impression of physics (Marshall and Linder, 2005). Content creation should be *research-based*. Textbook creators should support understanding with illustrations (Ogan-Bekiroglu, 2007), aim for high pedagogical quality (Marshall and Linder, 2005) and according to Zuza *et al.* (2020), content creation should involve design-based research, apply epistemological and psychological insights, apply the Vygotskian learning theory, assess new textbooks in the classroom, and consider students' learning paths to overcome preconceptions and comprehension difficulties.

To create *competence-oriented* textbooks, authors should include tasks for high-level skills (Takaoğlu, 2018; and Türk *et al.*, 2018), foster creativity (Klieger and Sherman, 2015), consider diverse cognitive domains according to Piaget (Barojas and Trigueros, 1991) and integrate thought experiments (Bancong and Song, 2018). To make textbooks *student-oriented*, creators should consider students' interests and needs (Takaoğlu, 2018), use a writing style and layout that catches students' attention (Ogan-Bekiroglu, 2007; and Mahardika *et al.*, 2020), write at the students' level (Fatoba, 2014; and Fitriah, 2020), provide meaningful content, and offer contents for motivation and review (Fitriah, 2020). Creators should make textbooks *interdisciplinary* by focusing on social, philosophical and historical aspects (Marshall and Linder, 2005), by integrating religion and local wisdom (Fitriah, 2020), as well as by integrating other skills such as new literacy and disaster literacy (Mufit *et al.*, 2020). According to Mufit *et al.*, new literature includes data literacy, technology literacy and human literacy. Researchers recommend textbooks to be made *stereotype-free*, especially with respect to gender, with authors considering gender differences (Artuso, 2017) while being gender-neutral (Gumilar and Amalia, 2020). Moreover, creators should make textbooks *context-based* by connecting scientific concepts to the students' world (Ogan-Bekiroglu, 2007; and Lous and Garcia, 2020). Finally, creators need to check that the textbook is *error-free* (Artuso, 2017).

Regarding form, researchers recommend textbook creation to be *multi-representational*, with authors offering diverse tasks (Takaoğlu, 2018), presenting concepts in multiple modes and including mathematical models and thought experiments (Rozina, 2006), using concept cartoons (Ogan-Bekiroglu, 2007) and involving other forms of representation (Handayani *et al.*, 2019; and Citra *et al.*, 2020). Creators should make textbooks *content-method-aligned* by offering abstracts and diagrams (Artuso, 2017), and by stating learning demands in line with the students' zone of potential development according to Vygotsky (Zuza *et al.*, 2020). Moreover, textbooks should be *clearly laid-out* with a systematic structure (Fitriah, 2020), with enough line spacing, and a list of units, index, and glossary (Ogan-Bekiroglu, 2007). Publishers are recommended to go *fully digital*, offering interactive online textbooks (Tesfaye and White, 2014).

Regarding actions, researchers recommend textbook creation to be based on *interaction*: Curriculum-makers should understand authors' and publishers' motives. Authors should be among the curriculum-makers (Ververs, 2016). Researchers should conduct peer review before textbook publication (Slisko, 2014). Authors should apply learning principles (Newton, 1984). Students should get information about the authors (Ogan-Bekiroglu, 2007). Software developers should work for e-book authors (Brajkovic, 2014). Most importantly, authors should consider teachers' expectations, students' needs and science researchers' knowledge (Duit *et al.*, 1992). Beyond that, they can learn from other good textbooks (Wiyanto *et al.*, 2020).

Researchers advise creators to include supplementary materials, especially materials that offer complementary contents (Newton, 1984), and that present data and graphs interactively (Rozina, 2006). To enable the *choice of textbooks,* authors should address diverse educational levels (Zuza *et al.*, 2020), and curriculum-makers should allow for diverse textbooks (Lous and Garcia, 2020). Following a *critical reading* of textbooks, teachers may even want to write their own textbooks (Handayani *et al.*, 2019). To *encourage textbook use*, textbook creators should place tasks throughout each textbook unit (Takaoğlu, 2018).

14.5.7.2 Recommendations for textbook use

Regarding content, researchers recommend textbook use to be *curriculum-aligned*, with the textbook serving as a curriculum guide (Aguiar and Garcia, 2019).

Regarding actions, researchers recommend textbook use to be based on a good *choice of textbooks*, with the teacher choosing the textbook (Trebien and Garcia, 2013). The teacher should use a textbook which addresses students' viewpoints (Podolefsky and Finkelstein, 2006), is possibly bilingual (Sipayung, 2020), has a simple writing style, includes illustrations, diagrams, local examples and activities, and is ideally available for free (Fatoba, 2014). Often, the textbook needs to be *supplemented*, with students and teachers using additional materials (Schultz, 1989; and Tesfaye and White, 2014), for example about the history and philosophy of science (Rodríguez and Niaz, 2004a), or materials to promote creativity (Klieger and Sherman, 2015), or improvised materials (Marshall and Linder, 2005), or digital resources (de Souza and Garcia, 2019). Textbook use should be *encouraged*, for example by the teacher giving tasks that require textbook use, also by grading pre-class reading (Heiner *et al.*, 2014), and by using textbooks that are at the students' level (Fatoba, 2014). When using textbooks, students should always *read critically* (Newton, 1984).

14.5.7.3 Recommendations for textbook research

Regarding content, researchers recommend textbook research to be *nature-of-science-oriented*, investigating the influence of physics textbooks on physics research (Artuso, 2017). Moreover, research should be about *stereotypes* in textbooks, especially checking for gender neutrality (Gumilar and Amalia, 2020).

Regarding form, researchers recommend that textbook research be focused on *fully digital* textbooks, comparing different software to create e-books (Brajkovic, 2014).

Regarding actions, researchers recommend textbook research to be based on *analyzing textbooks*, for example, using a particular instrument for analysis, dealing with different topics (Ververs, 2016), comparing tasks in textbook vs entry exams (Takaoğlu, 2018), identifying stylistic devices, studying the rhetorical model (Strube, 1989), checking if the history and philosophy of science is included (Rodríguez and Niaz, 2004a), investigating factors of translation quality (Sipayung, 2020), investigating the transition between cognitive domains (Barojas and Trigueros, 1991), or determining the readability level (Fatoba, 2014).

Textbook research should be based on *interaction*. Specifically, researchers should explore teachers' views on textbooks, explore how teaching is influenced by textbooks and curricula, study authors' intentions for textbook use, study textbook use (Ververs, 2016), consider the classroom and curriculum context (Trebien and Garcia, 2013), help authors in textbook writing (Strube, 1989), investigate to what extent authors know diverse physics methodologies (Rodríguez and Niaz, 2004a), investigate the impact of textbooks (Martins and Garcia, 2015), and even design a textbook (Artuso, 2017). Textbook research should include an *investigation of effects*, with researchers investigating, for example, how questions influence critical reading (Heiner *et al.*, 2014), how contextualized tasks affect learning (Slisko, 2014), how an application of the theory of cognitive domains influences students' cognitive ability and argumentation skills (Handayani *et al.*, 2019), or how communication and collaboration are promoted by textbooks (Wiyanto *et al.*, 2020).

14.6 DISCUSSION

With our literature analysis, we have answered the following question: Which expectations about physics textbooks have been found, and how? For this, we have analyzed all relevant sections of each paper: the introduction (focusing on the research gap and research question), methods, results, and discussion (focusing on recommendations for textbook creation, use, and research). Now, we will discuss our findings for each of these sections.

14.6.1 Discussion of the papers' research gaps

In the 39 analyzed papers, the section "research gap" contains a wide range of implicit expectations. In the eight empirical papers, the **research gaps were only related to content-based expectations**, namely, that textbooks should be more *student-oriented* (Marshall and Linder, 2005; Ververs, 2016; and Artuso, 2017), *course-matching* (Tesfaye and White, 2014; and Martins and Garcia, 2015), *competence-oriented* (Newton, 1984; and Podolefsky and Finkelstein, 2006), and *research-based* (Ogan-Bekiroglu, 2007). On a closer look (see chapter 5.3), we see that these content-related research gaps are physics-specific.

In three of the eight empirical papers (Marshall and Linder, 2005; Ververs, 2016; and Artuso, 2017), the authors point out that textbooks are not *student-oriented* enough. "Textbooks do not [...] fit the actual needs, abilities, and interests of students [...] but teachers' views of students' needs, abilities and interests" (Duit *et al.*, 1992, p. 107). As a countermeasure, some researchers are "in favor of including students in textbook research" (Knecht and Najvarová, 2010, p. 1).

In the normative and literature-based papers, other research gaps were pointed out. According to these 31 non-empirical papers, textbooks are not *curriculum-aligned*, not *multi-representational* enough, not *fully digital*, not *stereotype-free*, not *nature-of-science-orientated* enough, not *context-based* enough, and not *interdisciplinary* enough. These categories are not physics-specific but also represent problems in science education and generic education research. For example, the nature of science (NOS) is also an important topic in science TER, cf. Roseman *et al.* (1996). Arguably, some research gaps in PTER exist because—or although—**similar gaps have already been pointed out in science education and generic education research**.

14.6.2 Discussion of the papers' research questions

In the 39 analyzed papers, a wide range of textbook expectations was expressed in the section "research question.". Regarding educational levels, most of the analyzed papers are about secondary education, followed by tertiary and primary education. We find this **preference for secondary education** not only in physics education research but also in generic education research.

As expected, the expectations mentioned in the section "research question" of the eight empirical papers are about the textbook **content**, similar to the expectations of the section "research gap." Textbooks are expected to be *student-oriented* (Newton, 1984; Marshall and Linder, 2005; Podolefsky and Finkelstein, 2006; Heiner *et al.*, 2014; and Artuso, 2017), *competence-oriented* (Ogan-Bekiroglu, 2007), *course-matching* (Tesfaye and White, 2014), and *teacher-oriented* (Martins and Garcia, 2015). In each of the empirical papers, the research question falls into the same category as the research gap, except for two papers (Ogan-Bekiroglu, 2007; and Martins and Garcia, 2015). In these two papers, the research question is much more specific than the research gap, leading to a more specific category.

Five of the eight empirical papers contain questions about the use of textbooks (Newton, 1984; Podolefsky and Finkelstein, 2006; Tesfaye and White, 2014; and Ververs, 2016) and textbook evaluation (Ogan-Bekiroglu, 2007). All these research questions contain only *indirect statements* of expectations. Only three empirical papers mention *direct statements of expectations* in their research questions (Marshall and Linder, 2005; Martins and Garcia, 2015; and Artuso, 2017). Thus, in PTER, **explicit inquiries about expectations are even less frequent than implicit inquiries**.

Thus, research on explicit expectations about physics textbooks is almost non-existent.

Why is that? Let us find out by zooming out into textbook research in general.

Undoubtedly, textbooks are important (Ogan-Bekiroglu, 2007). Thus, textbook research should be abundant. However, globally, there is not much textbook research happening in general and even less in science (Bölsterli Bardy, 2014). For example, in the Taylor & Francis Group, the number of papers on science textbook research is low (Bölsterli Bardy, 2014, p. 16). Search results for the keyword "textbook" were five times fewer than for "computer," even though textbooks play a greater role in teaching than computers (Bölsterli Bardy, 2014, p. 16). Similarly, in a Brazilian review of physics textbook research, only 15 papers were found for the period between 2009 and 2017 (dos Santos *et al.*, 2019, p. 51). Thus, **textbook research is only a small field**.

How does textbooks expectations research (TER) fit into that field of textbook research? Generic textbook research can be divided into three areas: (1) Process-related research on the production, approval, selection, and use of textbooks, (2) product-related research about the quality of textbooks, and (3) performance-related research on the impact of textbooks on teachers and students (Mayer *et al.*, 2000, p. 5). 20 years ago, the focus of textbook research has expanded from product-related to process-related research (Horsley, 2002). For physics-specific textbook research, eight topics can be identified: (1) Constitution of the textbook, (2) environmental education, (3) experimentation, (4) science history, (5) paradidactic books, (6) problem solving, (7) imaging, (8) representations and didactic transposal (dos Santos *et al.*, 2019, p. 51). Looking at these areas of generic and physics-specific textbook research, we realize that **textbook expectations research is not a research area of its own**. This may be one of the reasons why PTER is rare.

Looking at the world map (Fig. 14.1), we see that **PTER is strongly limited in space and time**. We have found PTER only in a few countries. Interestingly, the dominance in PTER has clearly shifted from North America to Brazil around the year 2013 and from Brazil to Indonesia around the year 2020. Only a few English contributions have come from Europe, Africa, Australia, and Asia (except Indonesia). We only found links between the papers from Garcia (Trebien and Garcia, 2013; Martins and Garcia, 2015; Aguiar and Garcia, 2019; de Souza and Garcia, 2019; and Lous and Garcia, 2020). The other papers did not refer to each other.

Why have we not found more papers about PTER? Probably, **most of the relevant research has only been published in the local language**, which is rarely English. For example, in German-speaking countries, we found many authors only publishing in German. We see this tendency both for PTER and science TER (Bölsterli Bardy, 2014).

At least, **PTER** is growing. Whereas we have found no publication from before the Sputnik shock and only 5 papers from before the PISA shock, we have found 32 papers from the PISA shock and before Covid-19.

14.6.3 iscussion of the papers' methods

Most of the analyzed papers are normative or literature-based. Only 8 out of 39 papers are empirical. The preferred method of inquiry is to use a questionnaire, a convenient and reliable tool in physics education research. In accordance with the results from the section "research question," only three papers (Marshall and Linder, 2005; Martins and Garcia, 2015; and Artuso, 2017) perform **explicit inquiries about textbook expectations**. Although some papers have used both qualitative and quantitative methods of analysis, only qualitative methods have been asking about textbook expectations. This is well comprehensible, as you have to ask in an open format to empirically raise new textbook expectations (Bölsterli Bardy, 2014).

14.6.4 Discussion of the papers' results

In the results part of this review (see Sec. 14.5), the expectations mentioned in the section "results" are stated in detail. Here, the occurrence of content-related and form-related textbook expectations of the section "results" is summarized in Tables 14.6 and 14.7 for a better overview, showing which expectation holder mentioned which category for which part of a textbook.

Table 14.5
Categories of expectations related to actions related to textbooks.

Categories related to actions	Abbreviation	Definition
Supplement the textbook	ST	The book should be supplemented by other (analog or digital) media.
Analyze textbooks	AT	The content or form of a given textbook should be analyzed qualitatively or quantitatively by education researchers.
Interact	IA	Direct and indirect creators and users of textbooks should learn from each other to improve the creation and use of textbooks.
Investigate effects	IE	The effect of textbooks on students' learning should be studied.
Read critically	RC	A textbook should be read with a critical mind.
Choose textbook	CT	A textbook must be chosen by the teacher such that it is suitable for the course.
Encourage textbook use	ET	Student use of the textbook before, during, or after class must be encouraged by the teacher.

Table 14.6

Overview of the expectations about the textbook content mentioned in Sec. 14.5. C, curriculum-makers; T, teachers; S, students; R, researchers; A, authors; and P, publishers.

		Categories related to expectations about the content of a textbook									
		Course-matching	Nature-of-science-oriented	Competence-oriented	Stereotype-free	Curriculum-aligned	Inter-disciplinary	Student-oriented	Teacher-oriented	Context-based	Error-free
Structural elements of textbooks	Contents	T,S	C,T,S,R	C,T,S,R	R	T,S,R	S,R	C,T,S,R	…	C,S,R	T
	Keywords	…	R	…	…	…	…	…	…	…	…
	Main Text	S	C,R	T,R	…	S	…	T,S,R	…	T,R	S,R
	Highlighted text	…	…	T	…	…	…	…	…	…	…
	Figures	S	R	…	…	…	…	T	…	T,R	T,S,R
	Experiments	C,T,R	T,R	C,R	…	T	…	T,S,R	…	R	…
	Tasks	S	R	T,S,R	…	…	…	T	…	R	…
	Whole book	R	T,R	T,S	T	T	…	T,S,R	…	T	T,S,R
	Supplementary materials	T,S	C	T	…	…	…	…	…	…	…

Table 14.7
Overview of the expectations about the textbook form mentioned in the section “results.”

		Categories related to expectations about the form of a textbook					
		Multi-representational	Method-content-aligned	Clearly laid-out	External criteria	Fully digital	Research-based
Structural elements of textbooks	Contents	T,S,R	R	…	…	…	R
	Keywords	R	…	T	…	…	…
	Main Text	…	R	T	…	…	S,R
	Highlighted text	…	S	T	…	…	…
	Figures	T,S,R	T,R	T,R	…	…	T
	Experiments	T	R	…	…	…	…
	Tasks	R	R	…	…	…	…
	Whole book	S,R	T,S,R	T,S	T,S,R	R	T,R
	Supplementary materials	…	…	…	…	T,S	…

Overall, we observe a wide range of textbook expectations gained in the results of the analyzed articles. All categories of **content-related** textbook expectations can be found in the results section of the papers, except the category of being *teacher-oriented.* Similarly, all categories of **form-related** textbook expectations are present in the results section of the papers. However, the more interesting part is that not all expectation holders mention all categories (see Tables 14.6 and 14.7) and that the expectations only focus on some structural elements of a textbook.

Among the categories related to expectations about the content of textbooks, the following categories were found most widely across the structural elements of textbooks, see Table 14.6: *CB (be context-based), CO (be competence-oriented), NO (be nature-of-science-oriented), SO (be student-oriented)*, and *CM (be course-matching).* They match with prominent topics in contemporary physics education research and generic education research (Hattie, 2009; and Helmke, 2012). Apparently, **physics education researchers often transfer topics of their field into their own expectations about textbooks.** A good example is the analyzed article of Duit *et al.* (1992) in which the topics “girl-suited” science teaching, Science-Technology-Society, and constructivism were transferred to textbook expectations**.** However, without inquiring into teachers’ and students’ textbook expectations, we cannot know whether this transfer is adequate (Bölsterli Bardy, 2014).

Among the categories related to expectations about the form of textbooks, the following categories were found most widely across the structural elements of textbooks, see Table 14.7: *MR (be multi-representational)*, and *CM (be content-method-aligned).* Again, these reflect prominent topics of physics education research. Topics of generic education research, such as clear *layout (LO)* and being *fully* *digital (FD)*, are rarely mentioned. Presumably, **knowledge for textbook creation regarding layout and digitization is mostly adopted from generic education research.**

Curriculum-makers' expectations are purely content-related (see Tables 14.6 and 14.7). Curriculum-makers mention expectations about contents, main text, experiments, and supplementary materials (Table 14.6). Apparently, curriculum-makers translate their knowledge of curriculum development directly into textbook expectations without considering form-related aspects. Hence, it is crucial to involve different expectation holders in creating textbooks, each expectation holder contributing a unique perspective (cf. Bölsterli *et al.*, 2014).

Authors' and publishers' expectations have not been investigated in any of the analyzed papers (see Tables 14.6 and 14.7). However, in science education research, authors' expectations have been inquired (Bölsterli *et al.*, 2014), and publishers' expectations have been normatively stated (Duit *et al.*, 1992).

Researchers' expectations are expressed more abundantly than teachers' and students' expectations (see Tables 14.6 and 14.7). Evidently, researchers define their own expectations about textbooks, independent of the expectations of textbook users, cf. Ogan-Bekiroglu (2007); and Knecht and Najvarová (2010).

We see **no major difference between the textbook expectations of teachers' and researchers' expectations** (see Tables 14.6 and 14.7). However, in other studies, the expectations of science textbooks between teachers and researchers were quite different (Bölsterli *et al.*, 2014). In our review, the match between teachers' and researchers' expectations may be explained by the fact that most of these expectations are derived from a single study by Ogan-Bekiroglu (2007) in which teachers and researchers interacted, leading to some consensus.

Most notably, **teachers, researchers, and students agree that textbook contents should be both *curriculum-aligned* and *student-oriented*** (see Table 14.6)**.** However, these expectations are hard to fulfill simultaneously. One extreme case is the Karlsruhe physics course (Karlsruher Physikkurs, KPK) (Herrmann, 2000; and Herrmann, 2022). The goal of the KPK is to reconstruct scientific concepts to give students a modern and understandable perspective on physics (Herrmann and Job, 1996). The central idea is to consistently formulate densities and flows for mass, energy, momentum, angular momentum, electric charge, entropy, and amount of substance (Schmid, 1984). Although the KPK was tested and marketed in Germany in the 1990s and 2000s, it was denounced by the German physics society (DPG) in 2013 due to its radical reconstruction of scientific concepts. However, the KPK textbook has been accepted in Shanghai (China) and has become the basis for textbooks in two German high schools. Moreover, online textbooks for the KPK are available in Chinese, English, French, German, Italian, Russian, Spanish, and Swedish (Herrmann, 2022). Ultimately, it is the teacher who decides whether a given textbook is useful for students, and whether it can be aligned with the nation-specific curriculum.

14.6.5 Discussion of the papers' recommendations

In Table 14.8, we have summarized the recommendations gathered in the discussion sections of the analyzed papers. We see that recommendations for textbook creation and textbook research are related to content, form, and actions, whereas recommendations for textbook use are only related to content and actions (see Table 14.8).

The following action-related categories are mentioned both as recommendations for textbook creation and use, see Table 14.8: *supplement the textbook, read critically, choose textbook,* and *encourage textbook use.*

In both textbook research and textbook creation, **researchers have called for more interaction among expectation holders** (Table 14.8). It is important for researchers to interact with all expectation holders, as emphasized not only in physics education (Ogan-Bekiroglu, 2007) but also in science education (Bölsterli *et al.*, 2014) and generic education (Knecht and Najvarová, 2010). Experts in science TER have called for increased collaboration, especially in textbook creation (Bölsterli *et al.*, 2014).

Table 14.8
Recommendations made in the discussion section of the papers, cf. Tables 14.3–14.5.

	Content-related recommendations (cf. Table 14.3)	Form-related recommendations (cf. Table 14.4)	Action-related recommendations (cf. Table 14.5)
Recommendations for textbook creation	Nature-of-science-oriented	Research-based	Supplement the textbook
	Competence-oriented	Multi-representational	Interact
	Stereotype-free	Method-content-aligned	Read critically
	Interdisciplinary	Clearly laid-out	Choose textbook
	Student-oriented	Fully digital	Encourage textbook use
	Context-based		
	Error-free		
Recommendations for textbook use	Curriculum-aligned	…	Supplement the textbook
			Read critically
			Choose textbook
			Encourage textbook use
Recommendations for future physics textbook-research	Nature-of-science-oriented	Fully digital	Analyze textbooks
	Stereotype-free		Interact
	Student-oriented		Investigate effects

Overall, most recommendations are about textbook creation, see Table 14.8. Textbook researchers would like authors to implement their research results. For textbook use, researchers have only made a few recommendations. Three categories are not addressed at all in the recommendations: *be teacher-oriented, be course-matching* and *fulfill external criteria.* Apparently, **researchers are too distant from the textbook users**. In fact, in science TER, a significant gap between researchers' and teachers' views have been observed regarding the suitability of textbooks for daily use (Bölsterli *et al.*, 2014).

For future research, the authors of the 39 papers have made many recommendations, especially for textbook analysis and investigations into textbook effects, see Table 14.8 (bottom row). However, we have found **no recommendation to further explore expectations** about textbooks. Here again it becomes obvious that PTER is almost non-existent.

Ideally, the recommendations of one paper would become the research gap for the next paper. Unfortunately, we have found no such research paths. On the contrary, most papers stand for themselves, citing papers that are not specifically about expectations. The situation in science TER is similar. However, some research projects in science TER refer to each other, at least within one country [e.g., for the U.S.: Chiappetta *et al.* (1991) → Chiappetta *et al.* (2004) → Philipps (2006) → Lee (2007) (Chiappetta *et al.*, 2004; and modified 2006 and 2007)].

14.6.6 Limitations of the review

Our literature review has several limitations regarding the methods for literature search and inductive categorization.

For the literature search, we have only used Google Scholar, we have applied our keyword search only to the title, we have considered only English texts, and we have only used articles that are specifically about physics education, not science education.

For the inductive categorization of expectations, the constructed set of categories is only one of many possible sets. Moreover, we have subsumed each expectation under only one category (even if multiple categories might have matched). Finally, we have divided up the work of applying the categories so that only some of the expectations were categorized by all authors.

14.7 CONCLUSION AND OUTLOOK

In this chapter, we have reviewed expectations about physics textbooks held by curriculum-makers, authors, publishers, teachers, students, and researchers. Unfortunately, physics textbooks expectations research (PTER) is not a research area of its own, so research on textbook expectations is very rare, unsystematic, and mutually unrelated. Even among the papers that are about PTER, only a few of them are based on an explicit inquiry: Most are purely normative or literature-based.

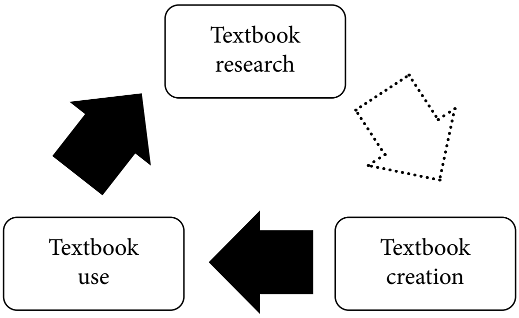


FIG. 14.4
Relationships between textbook research, creation, and use (black arrow: actual; dashed arrow: potential). Textbook research includes normative, literature-based and empirical expectations.

Taking together the contributions from all analyzed papers, we have compiled a long catalog of expectations. Still, this catalog is incomplete because not all expectation holders have been asked systematically about all structural elements of a physics textbook. In particular, the authors' and publishers' perspectives are unknown, and students' and teachers' expectations have rarely been investigated.

While there is a lot of literature about how physics textbooks *are* based on textbook analysis and evaluation, there is hardly any literature about how physics textbooks *should be*. It is self-evident that physics textbooks should serve the teacher and the student, but only a few research articles have identified specific characteristics that make a textbook useful. In the future, a more extended and targeted approach to physics textbook expectations research is needed.

Ideally, there should be a research cycle between textbook research, creation, and use (Fig. 14.4). However, there will always be a missing link between textbook research and textbook creation until there is more interaction among all expectation holders. It is time for physics education researchers to take textbook creators, and users' expectations more seriously and to enable communication between these expectation holders.

REFERENCES

Aguiar, C. F. and Garcia, N. M. D., *Researching Textbooks and Educational Media From Multiple Perspectives: Analysing the Texts, Studying Their Use, Determining Their Impact*, edited by Stig Toke Gissel (IARTEM, 2019), p. 207, https://www.ucviden.dk/ws/portalfiles/portal/124306891/IARTEM_2019_Proceedings.pdf#page=207.

Artuso, A. R., IARTEM e-J. **9**(2), 30–48 (2017).

Bancong, H. and Song., J., Jurnal Pendidikan IPA Indonesia **7**(1), 25–33 (2018).

Barojas, J. and Trigueros, M., Phys. Educ. **26**(3), 182 (1991).

Bölsterli Bardy, K., Am. J. Educ. Res. **3**(11), 1250–1254 (2015).

Bölsterli Bardy, K., Dissertation (Pädagogische Hochschule Heidelberg, 2014), Heidelberg, see http://nbn-resolving.de/urn:nbn:de:bsz:he76-opus-75385.

Bölsterli, K. *et al.*, *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning (Part 14)*, edited by C. P. Constantinou *et al.* (co-ed. Couso, D.; Louca, L.). European Science Education Research Association. (2014)

Bölsterli, K. *et al.*, *E-Book Proceedings of the ESERA 2011 Conference Science Learning and Citizenship*, edited by C. Bruguière *et al.* (European Science Education Research Association, 2012), pp. 14–18.

Brajkovic, M. (2014). Tools and methodologies for developing interactive electronic books: case study: A physics textbook for high school students. [Coursework] (Unpublished), see https://eprints.ucm.es/id/eprint/26508/.

Calinger, R. S. *et al.*, *Leonhard Euler's Letters to a German Princess* (Morgan & Claypool Publishers, 2019).

Chiappetta, E. L. *et al.*, J. Res. Sci. Tech. **28**, 713–725 (1991).

Chiappetta, E. L. *et al.*, (2004) (modified 2006 and 2007).: *Procedures for conducting content analysis of science textbooks* (Original from 2004, modified by M. Philips (2006), italicized modifications by Y. Lee (2007)). *Modified by M. Phillips (2006), italicized modifications by Y. Lee (2007), Department of Curriculum and Instruction, University of Houston*. Department of Curriculum & Instruction, University of Houston.

Cicchetti, D. V., Psychol. Assess. **6**(4), 284 (1994).
Citra, C. *et al.*, J. Phys. Conf. Ser. **1467**(1), 012029 (2020).
de Souza, J. L. L. and Garcia, T. M. F. B. (2019). *Researching Textbooks and Educational Media from Multiple Perspectives: Analysing the Texts, Studying Their Use, Determining their Impact*, 72.
dos Santos, T. A. *et al.*, *Researching Textbooks and Educational Media From Multiple Perspectives: Analysing the Texts, Studying Their Use, Determining Their Impact: 15th International Conference on Research on Textbooks and Educational Media*, edited by S. T. Gissel (Laeremiddel.dk, 2019), pp. 51–61.
Duit, R. *et al.*, Res. Sci. Educ. **22**(1), 106–113 (1992).
Fatoba, J. O., Int. J. Educ. Res. **2**(9), 41–50 (2014).
Fitriah, L., Jurnal Ilmiah Pendidikan Fisika **4**(1), 23–32 (2020).
Gumilar, S. and Amalia, I. F., Jurnal Keguruan dan Ilmu Tarbiyah **5**(2), 205–214 (2020).
Handayani, P. *et al.*, Phys. Conf. Ser. **1157**(3), 032039 (2019).
Haryanto, Z. and Syam, M., ISER (Indones. Sci. Educ. Res.) **2**(2), (2020).
Hattie, J., *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* (Routledge, 2009).
Haugsbakk, G., Educ. Inq. **4**(4), 23222 (2013).
Heiner, C. E. *et al.*, Am. J. Phys. **82**(10), 989–996 (2014).
Helmke, A., *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (4. überarbeitete Aufl., Schule weiterentwickeln–Unterricht verbessern. Orientierungsband) (Seelze: Klett-Kallmeyer, 2012).
Herrmann, F. (2022). Der Karlsruher Physikkurs, see http://www.physikdidaktik.uni-karlsruhe.de/.
Herrmann, F. and Job, G., Eur. J. Phys. **17**(4), 159 (1996).
Herrmann, F., Eur. J. Phys. **21**(1), 49 (2000).
Holovko, M. V., Проблеми сучасного підручника **16**, 69–80 (2016).
Horsley, M., *Learning and Educational Media: The Third IARTEM Volume*, edited by V. Meisalo *et al.* (Tarto University Press, 2002), pp. 11–29.
Kahveci, A., Int. J. Sci. Educ. **32**(11), 1495–1519 (2010).
Khoja, S. and Ventura, F., Libya. Mediterranean J. Educ. Stud. **2**(2), 119–129 (1997).
Klieger, A. and Sherman, G., Phys. Educ. **50**(3), 305 (2015).
Knecht, P. and Najvarová, V., J. Educ. Media, Memory, Soc. **2**(1), 1–16 (2010).
Koo, T. K. and Li, M. Y., J. Chiropr. Med. **15**(2), 155–163 (2016).
Lous, B. H. C. and Garcia, T. M. F. B., *Researching Textbooks and Educational Media From Multiple Perspectives: Analysing the Texts, Studying Their Use, Determining Their Impact* (IARTEM, 2020), p. 72.
Mahardika, I. K., J. Pengajar. MIPA **18**(2), 214–220 (2013).
Mahardika, I. K. *et al.*, J. Phys. Conf. Ser. **1465**(1), 012068 (2020).
Marshall, D. and Linder, C., Int. J. Sci. Educ. **27**(10), 1255–1268 (2005).
Martins, A. A. and Garcia, N. M. D., IARTEM e-j. **7**(1), 16–37 (2015).
Mayer, D. P. *et al.*, *Monitoring school quality—An Indicators Report*. National Center for Education Statistics, U.S. Department of Education (2000).
Merzyn, G., *Physikschulbücher, Physiklehrer und Physikunterricht Beiträge auf der Grundlage einer Befragung westdeutscher Physiklehrer (Eng. Physics textbooks, physics teachers, and physics education contributions based on a survey of West German physics teachers)*. Institut für die Pädagogik der Naturwissenschaften (1994).
Mikelskis, H. F., *Four Decades of Research in Science Education—From Curriculum Development to Quality Improvement*, edited by S. Mikelskis-Seifert *et al.* (Münster: Waxmann, 2008), pp. 57–65.
Mufit, F. *et al.*, J. Phys. Conf. Ser. **1481**(1), 012041 (2020).
Newton, D. P., Br. J. Educ. Technol. **15**(1), 43–51 (1984).
Ogan-Bekiroglu, F., J. Sci. Teach. Educ. **18**(4), 599–628 (2007).
Podolefsky, N. and Finkelstein, N., Phys. Teach. **44**(6), 338–342 (2006).
Pons, X., Eur. Educ. Res. J. **11**(2), 206–226 (2012).
Rodríguez, M. A. and Niaz, M., Instr. Sci. **32**(5), 357–386 (2004b).
Rodríguez, M. and Niaz, M., J. Sci. Educ. Technol. **13**, 409–424 (2004a).
Roseman, J. E. *et al.*, (1996). Identifying Curriculum Materials for Science Literacy: A Project 2061 Evaluation Tool: The report is based on a paper prepared for the Colloquium "Using the National Science Education Standards to Guide the Evaluation, Selection, and Adaptation of Instructional Materials," see http://www.project2061.org/publications/articles/roseman/roseman2.htm.
Rozina, I., Dissertation (University of Manitoba, 2006).
Schmid, G. B., Am. J. Phys. **52**(9), 794–799 (1984).
Schultz, F. H., Phys. Teach. **27**(4), 278–279 (1989).
Simon, J., *The Oxford Handbook of the History of Physics*, edited by J. Z. Buchwald and R. Fox (Oxford University Press, 2013).
Sipayung, K. T., Math. Hi. Lingua Cult. **14**(1), 79–85 (2020).
Slisko, J., The Eiffel tower as a context for word problems in textbooks for school mathematics and physics: why authors have a Licentia Poetica and what are possible consequences for students' learning and beliefs?. *Development (ICMT-2014)*, 433 (2014).
Strube, P., J. Res. Sci. Teach. **26**(4), 291–299 (1989).
Strube, P., Dissertation (University of Tasmania, 1985).
Takaoğlu, Z. B., Int. J. Assess. Tool. Educ. **5**(1), 58–72 (2018).

Tesfaye, C. L. and White, S., Nationwide Survey of High School Physics Teachers. Focus On. Statistical Research Center of the American Institute of Physics (2014).

Trebien, D. C. B. and Garcia, N. M. D., Danish University Colleges **18**, 108 (2013).

Türk, O. *et al.*, Eur. J. Educ. Stud. **5**, 42–55 (2018).

Valverde, G. A. *et al.*, *According to the Book. Using TIMSS to Investigation the Translation of Policy Into Practice Through the World of Textbooks* (Kluwer, Dordrecht, 2002).

Ververs, J. D., Master's thesis (Utrecht University, 2016).

Wiyanto, D. Y. *et al.*, Advances in Social Science, Education and Humanities Research, Volume 4. International Conference on Science and Education and Technology (ISET, 2019, 2020).

Zuza, K. *et al.*, Phys. Rev. Phys. Educ. Res. **16**, 2 (2020).

Zwahr, A., Auflage **24** (2006).

CHAPTER
15 TEXTBOOK AND CURRICULUM ALIGNMENT

Josip Slisko

Slisko, J., “Textbook and curriculum alignment,” in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 15-1–15-34.

15.1 INTRODUCTION

This chapter presents a literature review of different issues related to the complexities of alignments between physics textbooks and physics curricula understood in their broadest senses. Before a consideration of these complexities, it is useful to have initial views of basic common ideas about the central terms “curriculum” and “textbooks,” along with their relationships.

According to Tyler (1949), a school curriculum must contain four components:

1. Purposes of the school;
2. Educational experiences related to the school;
3. Organization of these experiences and
4. Evaluation of these experiences.

While in Tyler’s consideration “the school” was an institution responsible for curriculum design and implementation, Taba (1962) saw teachers as curriculum designers and the persons in charge to implement it. He presented a seven-step process of a curriculum making and enacting:

1. Diagnosis of needs of students for whom the curriculum is to be planned;
2. Formulation of objectives to be accomplished;
3. Selection of valid and significant contents;
4. Organization of contents taking into account the characteristics of the learners (cognitive maturity, academic achievements and interests);
5. Selection of learning experiences and instructional methods that would involve learners with the content;

https://doi.org/10.1063/9780735425514_015

6. Organization of learning activities into sequences that are determined by the content and learners' characteristics;
7. Evaluation and means of evaluation to determine which objectives have been accomplished.

English distinguished three types of curricula (formal, informal and hidden) and their three manifestation forms (written, taught and tested). He wrote a guide for teachers and administrators on how to deal with those six elements in developing, *aligning* and auditing a particular curriculum (English, 2000). English saw the textbooks as "curriculum surrogates," being unsatisfied with the fact that pupils use textbooks at least 75% of their classroom time:

> "These data suggest that the most important curriculum decision a district's official may make is not which curriculum to "develop" but which textbook to adopt." (English, 2000, pp. 15–16)

The two mentioned conceptions of curriculum design and implementation, institution - centered (state, school district, school) and teacher - centered (teacher who teaches a courses), represent two extreme points in a wide spectrum of situations that can occur in educational practices. Such varieties complicate the analysis of the relationship between curricula and textbooks.

According to Knight (2019), textbooks

a. include the body of knowledge that students are required to learn,
b. have all the relevant information in an organized structure that is sequenced, coherent and connective to study a discipline,
c. are peer-reviewed present accurate and complete information that is written by disciplinary experts who engage students in the culture and conventions of the discipline and
d. offer activities and quizzes relevant to learning and applying discipline knowledge.

Fey and Matthes (2018) explored quality criteria in different processes of textbook evaluations:

a. textbook production and development (authors, reviewers and publishers);
b. officially approving and recommending textbooks (government educational administration) and
c. selecting high-quality textbooks for educational purposes (schools).

Analyzing the textbook evaluation practices in different countries, Fey and Matthes (2018) found that in all of them, a relevant quality criterion of a textbook is its relationship with the corresponding curriculum. It is expressed with different wordings: "correspondence to curricular materials"; "compliance with curricula and subject-specific educational standards"; "alignment of aims, targets, and objectives with the curriculum"; "conformity to the curriculum," "compatibility with the curriculum" and "curricular congruence."

In more specific research domains of science education, an adequate relationship between the "curriculum" and "textbooks" is also recognized as important in their evaluation:

> “The textbook at each chapter clearly states the operational learning objectives/aims/goals presented in the national curriculum for a specific science subject. Also, competences that students should develop using the specific textbook chapter can be stated.
>
> Textbook content is derived from the learning goals stated in the national curriculum for the specific science subject, but not from the structure of the learning material. Textbooks help in achieving the learning goals and allow the students to achieve competences, both generic and subject-specific to science.” (Devetak and Vogrinc, 2013, p. 10)

The feature of that relationship is also among common research questions in mathematics textbook research:

> “How do textbooks reflect intended curriculum standards (if there are)?” (Fan, 2013)

15.2 THE PLACE AND ROLE OF TEXTBOOKS IN AN EXTENDED CONCEPTION OF CURRICULUM

To better understand the place and potential roles that textbooks play (or might play) in educational processes, a four-level view of curriculum (officially defined by educational authorities or supposed by authors) is very useful. It was developed by TIMSS (Trends in Mathematics and Science Study) as a framework for a big research project (Valverde *et al.*, 2002). The aim of the project was to define and compare basic features of mathematics and science textbooks in different countries (48 educational systems with over 400 textbooks).

The four-level view of curriculum has the following elements:

Intended Curriculum (IC)
Intentions, Aims & Goals;
Potentially Implemented Curriculum (PIC)
Textbooks and Other Organized Resource Materials;
Enacted Curriculum (EC)
Strategies, Practices & Activities;
Learned Curriculum (LC)
Knowledge: Ideas, Constructs & Schemes.

In contrast to the English’s view that the textbooks are “curriculum surrogates” (English, 2000), in this curricular conception textbook play crucial roles in connecting the IC with the EC and LC:

> “Textbooks are designed to translate the abstractions of curriculum policy into operations that teachers and students can carry out. They are intended as mediators between the intentions of

> the designers of curriculum policy and the teachers that provide instruction in classrooms." (Valverde *et al.*, 2002, p. 2)
>
> "Textbooks not only put forward the content students are to learn but they also advocate what students should be able to do with that content." (Valverde *et al.*, 2002, p. 125)
>
> "Textbooks … attempt to specify how classroom lessons can be structured. These specifications include an identification of the topics to be explored, their sequence, the activities that can be used in the exploration of the topic, and the behaviors that should be expected from students as part of this exploration. (Valverde *et al.*, 2002, p. 167)

The teachers involved in the TIMSS research project recognized the importance of the textbooks for designing and implementing their teaching:

> "Teachers reported in the TIMSS questionnaires that textbooks were a primary information source in deciding how to present content. Textbooks even had a major impact on decisions about what to teach and also on practical decisions about which instructional approach to follow and which exercises to use in class. Textbooks were the dominant source of information for planning what to teach in five of the 26 countries and the second most often cited by teachers in eight other countries." (Valverde *et al.*, 2002, p. 53)

The situations where "textbooks define curricula and not the other way around" were recognized before as something unfortunate by other authors. Yager (1983) described it by the following words:

> "Textbook determines contents and their sequence, classroom activities, homework and exams. Selection of the textbook is the most important decision for teachers."

Later Yager (1992) affirmed:

> "… Most teachers depend almost exclusively on textbooks to define their courses and to provide the activities for use in classrooms and teaching strategies to use in dealing with them. A good textbook is seen by teachers as one that makes life easy for the teacher user and keeps students engaged. Teachers love worksheets, directions for preparing laboratories, suggestions for quizzes, and chapter examinations.
>
> Ninety percent of all science teachers use a textbook (page by page) in excess of 90% of the time. For most students science becomes what is printed in textbooks and what is included on associated worksheets and in verification-type laboratories. Perhaps the most alarming fact is that by far the majority (75%) of existing teachers are quite satisfied with existing materials."

Norman L. Webb (1997) introduced the concept of alignment of "expectations" (stated in curriculum) and "assessments" (stated in applied exams) in mathematics and science education. Surprisingly, he didn't pay enough attention to instruction processes whose tasks are to give students all the experiences necessary for learning knowledge and skills that will be assessed.

Later developments (Squires, 2009; Polikoff and Porter, 2014; Polikoff *et al.*, 2015; and Seitz, 2017) extended the concept of alignment to include Enacted Curriculum (sometimes called Implemented or Taught Curriculum). Textbooks play, naturally, an important role in making possible curriculum enactment (Squires, 2009). Alignment curricular processes are quite complex, and consequently, the related studies are very demanding (Seitz, 2017; Ziebell and Clarke, 2018; and Raycroft and Flynn, 2020).

15.3 ALIGNMENT OF PHYSICS TEXTBOOKS AND CURRICULA IN PHYSICS EDUCATION RESEARCH: AN INITIAL CONTEXTUAL CONSIDERATION

In the framework of a four-level curriculum, possible research lines might be to consider the alignment of physics textbooks and curricula in at least two ways.

The first way would be to consider the physics textbooks as a dependent variable, posing the research question:

> How does the intended physics curriculum shape related physics textbooks as a potentially implemented curriculum?

It is obvious that such a research line implies that there exists an official intended physics curriculum and, additionally, that one or more physics textbooks were written with the aim to transform an intended curriculum into a potentially implemented curriculum.

That research line might not be very demanding if it is reduced to a factual or interpretative comparison between various elements of the contents listed in the curriculum and elaborated in textbooks. In such type of documental research, the goal is to find out the alignment level and extension of these elements. A review of this type of existing articles will be presented in a later section.

The second way would be to consider the physics textbooks as an independent variable, posing two general research questions:

> How do physics textbooks shape the enacted physics curriculum?
>
> How do physics textbooks shape learned curriculum?

This type of research might be very complicated because the textbooks are only one of a few possible independent variables that strongly influence how the intended curriculum is enacted by teachers and, consequently, which level of learning will students show on summative exams.

Teachers, with their professional preparation, motivation, dedication, and textbook selection, are surely principal actors in curriculum enactment in classroom or online settings. That important fact was explicitly stressed by Donahue (1993):

> "Curricula, no matter how reformed, were shaped by teachers' delivery to student."

Very few published articles have described these complicated processes in physics teaching and learning (for example, the relationship between elements of enacted and learned curriculum).

Some of them bring, at the first sight, puzzling results. For example, at course exams, the students who read physics textbook "often" outperform the students who read physics textbook "sometimes" but not those students who read textbooks "rarely" (French *et al.*, 2015). Nevertheless, the results indicate that the relationship between "reading textbook" (part of Enacted Curriculum) and the performance at "course exam" (part of Learned Curriculum) can't be causally determined only by taking into the account the "frequency of reading." At least, "quality of textbook reading" should be defined, measured, and added as a likely causal variable.

A more mindful approach to get a research-based alignment of curriculum and textbook (and/or other learning materials) can be found in some recent projects of physics curriculum developments. Namely, these projects develop, in an interactive and iterative process, both the curriculum (Brookes *et al.*, 2020) and the related physics textbook (Etkina *et al.*, 2019).

This fact makes an external analysis of their alignment not a very attractive research task. The reason is quite simple: when the same persons design, implement and carry out evaluative research of enacted and learned curriculum, and parallelly write and revise the textbook, it is unlikely that serious discrepancies could occur.

Articles focused on the alignment of physics textbooks and physics curricula are not very frequent in Physics Education Research (PER) and those articles, as will be shown below, are commonly written for the cases of educational systems outside the USA. It can be noticed by counting the presence of textbook-related articles in important review articles (McDermott and Redish, 1999; Thacker, 2003; and Doctor and Mestre, 2014) and in ComPADRE, a Digital Library of free online resource collections supporting faculty, students, and teachers in Physics and Astronomy education (compadre.org). For example, among 539 references listed by Doctor and Mestre (2014), only three of them have the term "textbook" in the title (a journal article, a Ph.D. thesis and a paper in conference proceedings).

The main reason for this fact is that in the United States, where the major development of PER happened and where the biggest number of active researchers work, an official national physics curriculum doesn't exist, both at high school and at college and university levels.

On the institutional level, the roles physics textbooks play in physics teaching and learning weren't in a deserved research focus. Since 1987, *the American Institute of Physics*, through its Statistical Research Center, has carried out a regular survey about the usage percentages of high school physics textbooks in different types of physics courses (Tesfaye and White, 2010). Although such percentages show interesting changes in the popularity of a particular textbook over time, there are no intentions to determine and understand the causal factors behind these changes.

The American Association of Physics Teachers (AAPT) neither planned nor carried out an organizational and massive research study related to the quality of physics textbooks used in the USA, that might be compared with the one undertook by TIMSS in an international context that was previously mentioned. Instead, only small-scale evaluative physics textbook reviews were published in *The Physics Teacher* (TPT).

The first review was produced in 1982 by a nine-member committee with Robert L. Lehrman as Chairman (Lehrman, 1982). The committee evaluated 14 physics textbooks for high school, using seven criteria that somehow agreed to be a common ground for their different “personal curricula” and visions of what a physics textbook should be:

1. Content (Accuracy and appropriateness of subject matter);
2. Level (Appropriateness of presentation for high school students);
3. Readability (The ease with which the book can be studied);
4. Appearance (Attractiveness of the book to the eye);
5. Science (Presentation of physics as a growing body of knowledge);
6. Social problems (Awareness of impact of physics on society);
7. Assignments (Adequacy of materials for additional works by students).

For the evaluation of each criterium, a four-point scale was used: poor, average, good and excellent. In the case of disagreement, an interval was indicated (for example, good to excellent).

The committee didn’t announce which textbook had the highest score and might be recommended by the evaluators, leaving the final choice to the teachers. The stated rational was (Lehrman, 1982):

> “We do not have any precise definition of what a physics text ought to do, and every teacher finds a different way of combining it with lectures, discovery experiments, verification experiments, problem sets and whatever other pedagogical techniques may be available. Further, each teacher sees a selected sample of high school students, and the book that is best for a highly motivated science-oriented candidate for a career in physics is probably altogether wrong for your average, run-of-the mill liberal arts type.”

The second review (Hubitz, 2001) was part of a grant from the David and Lucile Packard Foundation (Grant #1998-4248). The Foundation was especially concerned about errors in textbooks, but the grant was not just a routine “fact checking.” Its purpose was to review and critique middle school physical science textbooks and high school physics textbooks regarding to:

1. Scientific accuracy;
2. Adherence to a realistic portrayal of the scientific approach; and

3. Appropriateness and pedagogic effectiveness of the material for the grade for which it was presented.

John L. Hubitz, the principal researcher in the grant and who oversaw reviewing textbooks in physical sciences, presented an expanded list of the criteria used in the review (Hubitz, 2001):

1. Scientific Accuracy;
2. Adherence to an Accurate Portrayal of the Scientific Approach;
3. Appropriateness and Pedagogic Effectiveness of the Material;
4. Readability;
5. Attractiveness and Quality of Illustrations;
6. Laboratory Activities and Suggested Home Activities;
7. Exercises to Test Understanding; and
8. Resource Suggestions.

It is easy to see that this list is in resonance with the criteria of Lehrman (1982).

The report "Survey of high school physics texts" brought a qualitative review of 7 high school physics textbooks carried out by a 14-member Committee. Clifford Swartz, the editor of TPT and a committee member, summarized the results of the review (Swartz, 1999):

> "The most sophisticated, in terms of rigor of treatment and accuracy of presentation, is surely *PSSC Physics*. The most encyclopedic coverage is to be found in the new Holt book… *Physic-AL* is in a class by itself, with nonstandard sequence and treatment. It requires lab work matched to the text. *Conceptual Physics*… is still more appropriate for a preliminary course or for situations where students do not have to face an external exam in the standard topics… This is even more the case with the new and very different course called *Active Physics,* which is designed for students who will not be taking a standard physics course. There is no way to compare these texts in terms of one being "better" than another. Each is designed for a different audience with different goals."

The survey had two important sections. One section was a useful review of problematic details of content physics knowledge (Review Committee, 1999a). The other section (Review Committee, 1999b) dealt with the roles the textbooks potentially play in physics teaching and learning: Do students read physics texts? How do teachers use physics texts? How can students be persuaded to read the text? Should physics teachers encourage reading of texts?

The last paragraph of the Swartz summary was (Swartz, 1999):

> "The last time we reviewed high-school texts in *The Physics Teacher* was November 1982. Extrapolating, we can expect the next review in 2016. But will there be texts then?"

Expected linear extrapolation, predicting the time of the next review, didn't happen even until 2021! Additionally, the physics textbooks still exist and are an important part of physics education practice in the USA. Should the physics teaching community in the USA wait another generous grant to design and carry out some research projects to find out which are the best roles the physics textbooks could and should play in improving physics teaching and learning?

15.4 CURRICULUM-BASED RESEARCH OF PHYSICS TEXTBOOKS

As it was said before, strict curriculum-based research of physics textbooks to evaluate *expected alignment* is possible only in those countries where the Ministry of Education or a similar governmental institution publishes an official physics curriculum or physics study programs, playing the role of Intended Curriculum.

It is common that researchers in these countries incline to design and carry out explorations of textbook-related phenomena. Here are two illustrative examples.

The first example is related to the important (but rarely scientifically explored!) process of physics textbook development that is causally shaped by "physics curriculum" and authors' knowledge and skills (Lee and Lee, 2019). This is the case in which the textbooks were treated as a dependent variable.

The researchers were interested in the difficulties experienced by the authors during the development of physics textbooks. They formulated questionnaires and conducted interviews with seven authors with the aim to collect and analyzed the difficulties in the development of textbooks. 137 difficulties were found and classified into categories such as "overall composition," "level of content," "detailed composition & inquiry," and "illustration & photograph."

Some difficulties were related to the curriculum, such as "lack of understanding of curriculum," "inadequacy of inquiry," and "distinction of physics I and II." The greatest number of difficulties were found in the process of selecting the level and the scope of the content.

Based on the results of this study, the authors suggested "an improvement of curriculum development process," "a need to develop curriculum commentary," "an improvement of the textbook adoption system," and "a need for a textbook data support system."

The second example is related to students' interpretations and understandings of physics textbook illustrations (Park and Yoo, 2011). In this case, the textbook is an independent variable whose visual characteristics influence (together with other independent variables) students' knowledge (Learned Curriculum).

Authors affirm that physics textbook illustrations of standing waves in a pipe are not easy to understand because even a single illustration includes various representations such as the macrolevel, microlevels, concrete ways and abstract ways with line, arrows, and coloring. They wanted to investigate how high school students interpret and understand the illustrations of standing waves in a pipe.

The participants in the study were 145 high school students who learned standing waves in a pipe during the Physics 1 course.

Their recalled knowledge about standing waves was stimulated by one illustration, and their interpretations of the illustration components were surveyed using questionnaires. The responses were compared with the editor's intention for that illustration. To investigate tentatively the relations between students' interpretations and understandings of illustrations, they were asked to explain the movements of eight individual air particles in a pipe.

Wavelength formulas, an example of macro and abstract representations, were recalled by 67% of the participants. Only 6%, 9%, and 1% of the participants interpreted the lines, arrows, and coloring in the illustration in accord with the editor's intent.

Students showed a tendency to interpret illustrations on a macrolevel and in concrete ways rather than on a microlevel and in abstract ways. Various models were given when students explained the movements of individual particles. At most 2% of the participants seem to have scientific models, while most of the participants use mixed models.

When students constructed their own models for explaining the movements of individual particles, they seemed to consider one or two illustration components rather than all components. Some students did not seem to consider the given illustration at all, or they constructed their own models, neglecting the inconsistencies between the illustration and their own models.

These two important articles, that report the design and results of interesting and very needed textbook-related research, have only their abstracts written in English, while the rest are communicated in the Korean language. One can find additional examples of articles on the alignment between physics textbooks and curriculum written in Spanish (de Pro Bueno *et al.*, 2008; and de Pro Chereguini and de Pro Bueno, 2011) and very likely in other languages used in physics learning and teaching (Chinese, French, German, Italian, Japanese, Portuguese, Russian or Turkish).

This important fact is mentioned because in the literature review that follows, only articles published in English are selected, reviewed, and commented. This means that these studies represent only a very small fraction of global research activities related to the alignment of physics curricula and textbooks. It is an open question what the Physics Education Community should do in the future to stimulate the authors of research related to physics textbooks to publish their studies in English.

15.4.1 The selection of the reviewed studies

The search for studies that treat the alignment between "physics textbooks" and "physics curriculum" was carried out in Google Scholar. It is informative to learn how much Physics Education Research contributes to the general universe of studies that treat "textbooks," "curriculum" and their "alignment."

While the search term “textbooks” creates 2,490,000 hits, the term “physics textbooks” creates only 13,800 hits or 180 times less hits.

While the search term “curriculum” creates 3,910,000 hits, the term “physics curriculum” creates 16,200 hits or 240 times less hits.

While the search terms “textbooks” and “curriculum” create together 1,220,000 hits, the terms “physics textbooks” and “physics curriculum” create together only 830 hits or almost 1,470 times less hits.

Finally, while search terms “textbooks” + “curriculum” + “alignment” create together 105,000 hits, the terms “physics textbooks” + “physics curriculum” + “alignment” create together 260 hits or 400 times less hits.

These numbers show that Physics Education Research contributes relatively more to the general educational fields of “textbooks” and “curriculum” and far less to specific subfields of “textbooks and curriculum” and “alignment of textbooks and curriculum.”

The mere appearance of the terms “physics textbooks,” “physics curriculum” and “alignment” in the text of a study does not mean that it really considers their relationship as its research goal. Being so, the number of studies that were selected, reviewed and reported in this chapter is less than fifty.

15.4.2 Physics textbook studies carried out for a single country

A sample of studies related explicitly to textbook and curriculum alignment, due to the diversity of their contents and methodologies, will be presented in an alphabetic order.

15.4.2.1 China

The study of alignment of textbooks and new curriculum reforms in China dealt with “scientific inquires” (Li *et al.*, 2018). Reformed science curriculum from 7 to 9 grade, published in 2001, stated learning outcome:

> “Students should understand science knowledge through inquiry, obtain scientific skills, grasp scientific processes and methods, begin to understand the nature of science, form scientific attitudes, emotions, and values, and develop their innovative minds and practical abilities”

The revised 2011 science curriculum stressed even more the importance of inquiry in science education.

To determine whether the characteristics of scientific inquiry activities of textbooks in China satisfy the requirements of China’s new curriculum reforms, the researchers adopted a content analysis method to analyze scientific inquiry activities in five regional junior middle school physics textbooks (for grade 8) authorized by the Chinese Ministry of Education.

According to the researchers, to cultivate students' scientific inquiry abilities and scientific reasoning skills, the inquiry tasks should satisfy two standards. The first is "openness," meaning that scientific inquiry activities should have a lower degree of teacher involvement and a higher degree of student autonomy, meaning that they are open rather than following a recipe. Considering that openness is only one characteristic of authentic scientific inquiry, the researchers defined the second standard "operationality of cognitive processes," that is even more important to develop students' scientific inquiry skills and scientific reasoning skills. Scientific inquiry tasks should reflect all characteristics of cognitive processes that characterize authentic scientific inquiry.

The results show that the inquiry activities in these five physics textbooks do not meet the requirements of authentic scientific inquiry, regardless of the degree of openness or operationality of cognitive processes. Being so, analyzed textbooks are not conducive to developing students' scientific inquiry and scientific reasoning skills.

The researchers suggest that physics textbook authors should learn more about the latest insights from the world of academia and develop some new authentic scientific inquiry tasks to update textbooks in a prompt manner to improve the students' abilities of scientific inquiry. The textbook reviewers admitted that the reported research not only provided them with specific findings but also provided them with a method to evaluate textbooks (Li *et al.*, 2018).

15.4.2.2 Ethiopia

An interesting quantitative textbook study was carried out in Ethiopia (Zewdie, 2014). The article first presented what the curriculum has pointed out about physics learning:

> "The learning of physics enables students/learners to understand the physical world, to carryout observations and experiments related with physical events and phenomena, to enhance interest in nature."

After that, the author analyzed official physics textbooks for Grades 7 and 8 collecting comments from teachers during handbook familiarization. Attention was paid to "students' involvement" in the content elements: learning objectives, activities, figures and diagrams, text narratives, unit summaries, and end-of-unit exercises.

The results revealed that learning objectives were stated in all the units in the textbooks. In addition, the figures, diagrams, other drawings, and points that demand emphasis were put in attractive colors that might call students to read.

However, there were aspects of the textbooks with serious limitations: (a) emphasis on memorization of facts, explanations, and principles; (b) the activities had immediate answers in the textbooks; (c) figures and diagrams concentrated mainly for illustrative purposes; (d) review questions and problems demanded simple memory; and (e) mere mathematical calculations based on previous learning of formulas.

As such limitations contradict curriculum intentions for physics learning, in the conclusion, Zewdie gave recommendations for further improvements and revisions.

15.4.2.3 Nigeria

In the case of Nigeria, two studies related to the physics curriculum and textbooks called attention.

In the first study, Okoronka and Adeoye (2011) evaluated the presence of History and Philosophy of Science (HPS) in the contents of the Nigerian Physics Curriculum and five corresponding physics textbooks. The research questions were:

1. What are the specifications (topics, concepts, content, and activities) of HPS in the Nigerian physics curriculum and their adequacy?
2. What is the most frequently specified and used strategy of HPS in the Nigerian physics curriculum and textbooks?
3. What is the average number/percentage of HPS -related topics/contents in the curriculum?
4. In what context(s) are the HPS related content used in the textbooks?
5. Is the use of the HPS in Nigerian physics textbooks adequate?

The main results were as follows:

Only 17.4% of the topics in the curriculum contained issues concerned with HPS issues, among which are the models of the atom.

Only one textbook out of five reviewed dealt measurably well with the HPS topics as contained in the curriculum.

The most frequently specified strategy inferred from the curriculum is the historical model.

The physics textbook, that is most popular among students and teachers, has almost nothing in it to justify its attainment of HPS goals, while the rest of the physics textbooks dealt with HPS matters implicitly.

The general conclusion of the study was (Okoronka and Adeoye, 2011, p. 81):

> "… The percentage of HPS topics specified in the curriculum and treated in physics textbooks is considered grossly inadequate. Both the use of the HPS contents and the context of their use are inadequate. Therefore, it is suggested that a systematic program built around athe historical model, analogue strategy, classical experiments, and Vignettes be developed and implemented for the attainment of the aims and goals of HPS in elucidating the nature of science starting from our secondary schools."

The second study (Chukwunenye *et al.*, 2018) investigated the adequacy of three physics textbooks in use in Senior Secondary Schools in Nigeria (for the case of Owerri municipal, Imo state). Research questions were:

1. To what extent does the content of physics textbooks in use in senior secondary schools reflect the content of the new physics curriculum?
2. How adequate are the learning activities of the physics textbooks in senior secondary schools?
3. How adequate are the study questions contained in the physics textbooks for senior secondary school physics?
4. To what extent do the illustrations contained in the physics textbooks adequately represent the content of the new physics curriculum?
5. What are the teachers' perceptions of the physics textbooks in terms of compliance with the new physics curriculum?

Three textbooks (out of five recommended by the Ministry of Education) were evaluated by twenty-six physics teachers from ten public secondary schools.

The results revealed that all the textbooks were adequate in terms of content and study questions. Nevertheless, inadequacies could be detected in some of the texts in the areas of learning activities, illustrations, and teacher perception.

This simply implies that no single textbook has completely met the requirements of the new physics curriculum. Due to it, the authors recommended that teachers should not adhere to a particular textbook but rather should expose their students to a variety of textbooks depending on the goal or aim of the lessons.

15.4.2.4 Singapore

The curricular study in Singapore (Caleon and Subramaniam, 2010) explored the "learned curriculum" in the content domain of the properties and propagation of mechanical waves. To better evaluate this curricular aspect, the author considered what related to that domain was stated in "potentially implemented curriculum" (two standard physics textbooks) and "intended curriculum" (physics syllabus for grades 9 and 10).

This study reports on the development and application of an original four-tier multiple-choice (4TMC) diagnostic instrument. It is an enhanced version of the two-tier multiple-choice (2TMC) test. As in 2TMC tests, its answer and reason tiers measure students' content knowledge and explanatory knowledge, respectively. The two additional tiers measure the level of confidence of students in the correctness of their chosen options for the answer and reason tiers respectively. The 4TMC diagnostic test focused on the properties and propagation of mechanical waves. The final 12-item version of the diagnostic test was named 4WADI (Four-tier Wave Diagnostic Instrument).

4WADI was administered to 598 upper secondary students from six mainstream government co-educational schools in Singapore after they had experienced at least 6 h (spread over 3 to 4 weeks) of formal instruction on the properties and nature of waves.

Most of the respondents had an inadequate grasp of the topics tested. Mean scores and mean confidence associated with the answer tier were higher than those associated with the reason tier. The students tended to be poorly discriminating between what they know and what they do not know. Familiarity with the topic tested was associated with a greater percentage of students giving correct answers, higher confidence, and better discrimination quotient. Nine genuine alternative conceptions (which were expressed with moderate levels of confidence by students) were identified.

15.4.2.5 Slovenia

Among the objectives of the Slovenian National high school physics curriculum is the development of students' "complex thinking skills": thoughtful observation, reasoning, generalization, explanation and evaluation, modeling and autonomous problem solving (Planinšič *et al.*, 2008). Three studies were carried out to verify the alignment between one element of "complex thinking skills" (modeling) and authorized physics textbooks.

The first study (Forjan *et al.*, 2014) describes the results of analysis of the content of the *two most commonly used Slovenian secondary school physics textbooks*, from the standpoint of the presentation of modeling stages in physics problem solving: Simplifications of the problem (Conceptual model building); Visualization based on models of objects and interactions, Formulation of the mathematical model, Analysis of the mathematical model leading to the model results (solutions) and Validation of the model results.

The analysis focused on the extent to which these stages of modeling process were presented by the authors in the solved examples of physics problems.

The results show that the conceptual model building stage was very poorly presented. In most cases, the emphasis was on the stage of model analysis, i.e., solving mathematical equations. The validation stage of the modeling process was hardly detected in solved examples. This means that students would not be able to recognize the value of the experiment when assessing the relevance and accuracy of the developed models.

The second study (Forjan and Sliško, 2014) focused on the process of simplification in problem solving in the physics domains of thermodynamics, electricity, and waves. The presence of a clearly stated process of simplification was analyzed in solved problems in three most used Slovenian high school physics textbooks.

The rationale for the study was the fact that understanding simplifications is crucial for the modeling of physical systems since one doesn't consider irrelevant properties and minor effects and is focused on the most important characteristics of the systems and the processes. In high school physics, simplified and idealized models play a fundamental role in learning physics concepts and laws. Therefore, it is very important that textbooks present them carefully.

The results show that in all three textbooks, more than a third of analyzed simplifications are not properly presented and clarified. In addition, almost no explicit comment on assumptions and approximations used in the solved problems was found.

The third study (Forjan and Sliško, 2017) was an extension of the second one for the case of solved problems in mechanics. The review showed that in two textbooks, more than a half of the analyzed simplifications are not properly presented.

Further review was directed toward curriculum objectives (learning outcomes) that are directly related to the assumptions of limited validity of certain physical principles, which is the result of simplifications and idealizations of physical systems. In the curriculum, only two such objectives were found:

- students are aware of the limited validity of Hooke 's law,
- students are aware that the term $\Delta Ep = mg\Delta h$ has limited validity when moving away from the Earth.

In the content objective "Ohm's law and the definition of resistance," there is the comment "students know that Ohm's law does not apply to all conductors."

Based on this analysis of the physics textbook and curriculum, one can conclude that insufficient emphasis is given to the processes of simplifications or idealizations that are necessary and important for the theoretical treatment of physical processes and systems.

15.4.2.6 South Korea

The Special Theory of Relativity (STR) became a part of the South Korean physics curriculum in 2012. It happened in the following way (Gim, 2016):

The Ministry of Education of the South Korean government appointed some educational experts in physics and science education to define learning ingredients. These experts then wrote the learning standards that should be achieved by high school students. After these previous steps, authors were called to start to write textbooks. In the case of the STR, the achievement standard was to explain the fundamental *principles* of the STR (the constancy of the speed of light, time dilation, length contraction, simultaneity and the inter-convertibility of rest mass and energy). In South Korea, only two high school physics textbooks cover the STR.

Gim (2016) examined the portrayal of STR in two physics textbooks and discussed an alternative method to solve the analyzed problems.

This examination of how these textbooks present the theory has revealed two main flaws:

First, the textbooks' contents present historically fallacious backgrounds regarding the origin of this theory because of a blind dependence on popular undergraduate textbooks, which ignore the revolutionary aspects of the theory in physics.

And second, the current ingredients of teaching this theory are simply enumerated and conceptually confused. Consequently, the students are not provided with good opportunities to develop critical capacities for evaluating scientific theories.

Gim claims that the history of science contributes to understanding not only the origins but also the two principles of STR.

Second, in addition to this claim, Gim argues that we should distinguish not only hypotheses from principles but also phenomena from theoretical consequences and evidence. Finally, Gim suggests an alternative way in which theory testing occurs in the process of evaluation among competitive theories based on data and not in the simple relation between a hypothesis and evidence.

Inadequate historic, epistemological, and didactic approaches to teaching STR were also found in the most used physics textbooks in Argentina (Arriassecq and Greca, 2007).

15.4.2.7 Türkiye

In the case of Türkiye, two studies related to physics textbooks are worthy of mention.

In the first one, Bahçıvan and Eraslan (2011) critically analyzed the chapter on modern physics in the textbook for 10th grade elaborated by Türkiye's National Ministry of Education regarding scientific literacy, content appropriateness and instructional approach.

Although the authors consider that the textbook is compatible with the learning goals of the national curriculum, this compatibility is not complete. Namely, the textbook does not, but should, warn the students about common misconceptions in modern physics. In addition, according to the authors, the textbook content itself might lead students into some of these misconceptions. Classification of the scientific knowledge was overused in the textbook and students may think that the laws of mechanical physics and modern physics are completely different from each other.

Ogan-Bekiroglu (2007) designed the study with an ambitious goal: Developing an instrument to identify the characteristics of high school physics textbooks and examining how appropriate the currently used textbooks were for teaching and learning physics. She and her pre-service physics teachers reviewed the previously published literature on the quality of textbooks and, in a multi-step process, created an evaluation instrument for physics textbooks. Among the seven considered categories (or evaluation domains), the four most important were:

Content

Coherence with curriculum, scientific accuracy, appropriateness with the level of students, emphasis on scientific inquiry, level of questions, connection to the real word, and importance of the interaction between science, technology, and society.

> *Explanation and language*
>
> Structure and entity of text, explanation level, conceptual density, prevention of alternative conceptions, and appropriateness level of technical words for the grade level.
>
> *Activities*
>
> Relevance for the topic, appropriateness for the cognitive capabilities of the students, and emphasis on scientific process skills.
>
> *Instructional support*
>
> Providing a sense of purpose, taking account of student ideas, engaging students with relevant phenomena, developing, and using scientific ideas, promoting student thinking about phenomena, experiences, and knowledge, assessing progress, and enhancing the science learning environment.

In the second part of the study, pre-service physics teachers used the instrument to evaluate 11 physics textbooks approved by the Turkish Ministry of Education. Each textbook was graded by a team of four pre-service teachers. General grades for each category were 5 (very good); 4 (good); 3 (adequate) and 2 (weak).

None of the textbooks was graded as "very good" for any category. On average, for 11 textbooks, the categories "explanation" and "activities" were graded between "adequate" and "good." Nevertheless, average grades for the categories "content" and "instructional support" were between "weak" and "adequate." No details were given about which subcategory contributed to such unsatisfactory-grade values.

Some findings regarding basic physics concepts are disturbing: no textbook gives a correct definition of "electric current" or the discussion of weightlessness in two textbooks might lead students to conclude that there is no gravity above the atmosphere.

In the case of experiments, the students are told what should be observed and what conclusions should be reached. In addition, no safety rules are given.

The conclusion of the study is very strong (Ogan-Bekiroglu, 2007):

> "Unfortunately, the textbooks approved by the Ministry do not meet the criteria supporting the effective physics teaching and learning. It may be difficult for teachers to use these approved textbooks as means of promoting student inquiry; rather, they can use them as means of imparting factual knowledge."

As the Ministry of Education should, in principle, only approve the textbooks that are well aligned with the official curriculum, these negative results implicitly mean that the curricular framework does not reflect an adequate view of effective physics teaching and learning (as stated by subcategories of "instructional support").

15.5 ALIGNMENT OF PHYSICS TEXTBOOKS WITH META-CURRICULAR FRAMEWORKS

Meta-curricular frameworks are all those national or international documents that explicitly suggest what are knowledge and skill objectives that the pupils and students should be able to learn through educational programs. The list of knowledge and skills can be further stressed via important international evaluations such as PISA or TIMMS.

The most influential of such frameworks were American projects related to "science literacy." The starting point of the "*Project 2061*," launched by the American Association for Advancement of Science, was the critique related to the disturbing absence of "science literacy" in teaching and learning science:

> "The present science textbooks and methods of instruction, far from helping, often actually impede progress toward *science literacy*. They emphasize the learning of answers more than the exploration of questions, memory at the expense of critical thought, bits and pieces of information instead of understandings in context, recitation over argument, reading in lieu of doing. They fail to encourage students to work together, to share ideas and information freely with each other, or to use modern instruments to extend their intellectual capabilities." (Rutherford and Ahlgren, 1990, p. xvi)

"Science literacy" was related to:

> "…understanding some of the key concepts and principles of science; having a capacity for scientific ways of thinking; knowing that science, mathematics, and technology is human enterprises, and knowing what that implies about their strengths and limitations; and being able to use scientific knowledge and ways of thinking for personal and social purposes." (Rutherford and Ahlgren, 1990, pp. xvii–xviii)

It was stressed that the Project 2061 was neither a curriculum nor a textbook sketch:

> "The report deals only with *learning goals*—what students should remember, understand, and be able to do after they have left school as a residue of their total school experience—and not with how to organize the curriculum to achieve them. Neither is the presentation of recommendations meant to instruct the reader as a text does." (Rutherford and Ahlgren, 1990, pp. xx–xxi)

Scientifically literate persons should have knowledge about the way science works (scientific world view, scientific methods of inquiry, and the nature of the scientific enterprise).

The next contribution to the elaboration of "science literacy" was given in the "National Science Education Standards" (National Research Council, 1996) by making explicit what science and science learning are:

> “LEARNING SCIENCE IS AN ACTIVE PROCESS. Learning science is something students do, not something that is done to them. In learning science, students describe objects and events, ask questions, acquire knowledge, construct explanations of natural phenomena, test those explanations in many different ways, and communicate their ideas to others.” (National Research Council, 1996, p. 20)
>
> “Science is a way of knowing that is characterized by empirical criteria, logical argument, and skeptical review. Students should develop an understanding of what science is, what science is not, what science can and cannot do, and how science contributes to culture.” (National Research Council, 1996, p. 21)

The movement toward “science literacy” culminated in the book “A framework for K-12 science education: Practices, crosscutting concepts, and core ideas,” published in 2012 (National Research Council, 2012). Science education should make it possible for students to have productive learning experiences with the following scientific practices (National Research Council, 2012, p. 3):

1. Asking questions;
2. Developing and using models;
3. Planning and carrying out investigations;
4. Analyzing and interpreting data;
5. Using mathematics and computational thinking;
6. Constructing explanations;
7. Engaging in argument from evidence;
8. Obtaining evaluates and communicating information.

All these practices, taken together, help students to understand the “nature of science,” or, in other words, what science is and how it is done by scientists.

15.5.1 Nature of science in physics textbooks

A meta-curricular approach or framework was present in those studies that focused on the quantity and quality of the presence of the “nature of science” (NoS) in physics textbooks in different countries. As it was done before, due to differences in the selection of some aspects of NoS and methodologies, the studies will be presented in an alphabetic order. The list of countries in which these types of studies were conducted was determined by the interests of researchers and should not be equal to the previous list.

15.5.1.1 Finland and United States of America

In this rare comparative study, Park and Lavonen (2013) compared two high school physics textbooks used in Finland and the United States of America to examine how the curriculum is in alignment with the reform standards using the questioning style and level of inquiry activities, which are key components of the National Science Education Standards (NSES).

Two textbooks from two countries were chosen for analysis in this study: Physica (meaning physics in Greek) of Finland and Active Physics of the United States' high school physics. Both textbooks are products of reform efforts in science education.

In 2003, Finland undertook a major change in the curriculum at the national level, which produced the "National Core Curriculum," whereas science education in the United States went through a reform after 1996 when NSES was published.

Physica of Finland was developed as a high school physics textbook based on the National Core Curriculum for Science Education, considering its learning aims for upper secondary physics and short descriptions of core content. *Active Physics* of the U.S. was developed based on the NSES in the absence of a traditional curriculum.

Two major factors of Analysis Form were employed to analyze the selected samples of Active Physics and Physica: (a) general features and (b) level of openness of inquiry activities. The general features the authors adopted for analysis included the number of units, total pages, topics, and questions and sentences in each textbook.

In summary, both the U.S. textbook *Active Physics* and the Finish textbook *Physica* have more than 1,000 pages. Regarding the general features of the textbooks, *Active Physics* contains more laboratory activities and fewer chapters than Physica. While *Active Physics* has more experiential questions, the two textbooks have a similar pattern with regard to non-experiential questions, the majority of which involve direct-information and an open-ended question. According to the science and learning process scheme, the observing, experimenting, and inferring question types were common in both textbooks.

The authors concluded that, at least from the perspective of inquiry science and questioning type, both textbooks consistently aligned with the NSES visions, playing the role of a meta-curricular framework.

15.5.1.2 Indonesia

Ardwiyanti and collaborators (2021) assessed the representation of NoS quantitatively and qualitatively in two Indonesian XII grade high school physics textbooks in their chapters dealing with electromagnetic radiation, special relativity theory, and quantum phenomena The quantitative representation of NoS was described by the percentage of appearance of each element of NoS, while the delivery approach was used to describe NOS qualitatively. The results showed that the empirical and scientific theories were addressed with the highest percentage of appearance, while the social and cultural embeddedness of science and scientific methods were poorly addressed. All elements of NOS were represented implicitly. The researchers expect that their findings might be able to stimulate the improvement of the representation of NoS in science textbooks in Indonesia.

15.5.1.3 South Korea

Park *et al.* (2019) analyzed five new South Korean high school physics textbooks' conceptualizations and representations of NoS, particularly as reflected in their general relativity theory section. The results indicate that textbook references to NoS are concentrated on aspects related to scientific knowledge, scientific practice, scientific methods, and professional activities of scientists, whereas the characteristics of science as a social-institutional system are underrepresented.

15.5.1.4 United States of America

The most ambitious and influential longitudinal study on the incorporation of "nature of science" (NoS) in biology and physics textbooks was carried out in the United States of America (Abd-El-Khalick *et al.*, 2017).

This study assessed the (i) ways in which, and extent to which several aspects of nature of science (NOS) are represented in 16 high school biology and 18 physics textbooks in the United States (U.S.) and (ii) extent to which these representations have changed over the course of several decades.

NoS aspects included the empirical, tentative, inferential, creative, theory-laden, and social NoS; the myth of "The Scientific Method"; nature of theories and laws; and social and cultural embeddedness of science.

Textbooks were scored for the accuracy, way, and the extent (in textbook pages) the target NoS aspects were represented. Analyses indicated that, on average, only less than 2.5% of the analyzed textbook pages were dedicated to addressing NoS constructs. Overall, representations of NOS in the textbooks did not differ by content area, were discernibly less than favorable and did not improve substantially over the past several decades.

The study concludes that these trends are incommensurate with the emphasis placed in USA reform efforts on helping precollege students develop informed NOS conceptions.

15.5.1.5 Vietnam

Thao-Do and Yuenyong (2013) declared in their study that the term "nature of science" is foreign to Vietnamese teachers and students, although most of them have some certain understanding of common aspects of the nature of science (NoS) through implicitly teaching and learning science. Through a review of the Vietnamese physics textbooks, they found that along with the subject knowledge, there were several parts that mentioned the history of science. Therefore, Thao-Do and Yuenyong suggested several ways to link the historical information with some aspects of NoS explicitly in teaching heat, stressing the critical issue of energy degradation.

15.5.2 Explanations, thought experiments and use of analogies in physics textbooks

Specific scientific processes (explanations, thought experiments and use of analogies) illustrate a further *meta-curricular framework* defined by the Nature of Science. To understand these processes better, students should be informed about them or, even better, they should be given multiple opportunities to practice them. It is natural to expect that physics textbooks should provide elaborate examples and tasks to satisfy both learning needs.

Valentzas and Halkia (2018) analyzed the structure of scientific explanations included in three physics textbooks of upper secondary schools in Greece, whose contents correspond to the mandatory Physics curriculum. Later, they compared what they found with explanations in four internationally known American physics textbooks for university levels authored by Halliday *et al.*, (2008); Ohanian (1989); Serway (1990); and Young (1992). The aim was to trace whether the Greek authors followed similar reasoning in constructing scientific explanations as American authors.

In the scientific explanations that were mentioned for specific phenomena, the explanandum was a logical consequence of the explanans, which in all cases included at least one scientific law (and/or principle, model, or rule) previously presented, as well as statements concerning a specific case or specific conditions. The same structure was also recognized in most of the cases in which the textbook authors explained regularities (i.e., laws, rules) as consequences of one or more general laws or principles of physics. Similar logical structures in considered explanations were found in American physics textbooks. The study authors propose the following use of their findings:

> “Excerpts of scientific explanations from the textbooks could be given to students, who could then be asked to identify the explanandum, the set of sentences that constitute the explanans, and the reasoning that leads deductively to the explanandum from the explanans. A proposed future study would be the evaluation of such a procedure for the improvement of students’ ability in constructing scientific explanations of phenomena.” (Valentzas and Halkia, 2018)

Gilbert and Reiner (2002) explored the potential and actual roles of thought experiments (TE) in three typical physics textbooks, two of which were written by American authors (Hewitt and Ohanian) and one by a British author (Breithaupt). The authors first discussed the relationship between TE and the experiments supposed to be performed. Later, they addressed the potential uses of the various types of TEs in bringing about conceptual development and as a complement to conventional practical work.

The analysis of the textbooks found that the expected potential uses were not realized. Instead, the authors integrated the elements of TE with other pedagogic devices into what they termed “thought simulations." In these, the behavior of a phenomenon is illustrated rather than predicted and tested, theory is assumed and embedded rather than being tentative and emergent, and the outcome was assumed rather than anticipated.

In his well-known book "*Thinking physics is gedanken physics*," Epstein (1995) designed a number of interesting conceptual problems which are, in fact mini "thought experiments," being hard to actually perform. Some of these conceptual problems were overtaken and included in physics textbooks. Bancong and Song (2020) examined how 12 Indonesian pre-service and in-service physics teachers were solving these problems. They were interested in observing the students' solving processes while "doing thought experiments." These processes were classified as the following: prediction, verification, and explanation. Contrary to the scientific practice in which validation is theory-based, students validated their "thought experiments" on their experiences.

Yener (2012) scrutinized the types of analogies used in four physics textbooks in Turkish high schools. His research goal was to find out how these analogies are structured and presented. Analogies detected in physics textbooks were classified into a few categories: Analogical Relationship, Presentational Format, Condition of Subject Matter, Position in Text, Level of Enrichment, Pre-Topic Orientation and Limitations. Yener detected a total of 50 analogies. It was determined that these were mostly configured as functional, verbal, concrete-abstract, embedded activator, and simple analogies.

It was found that these analogies were used more for abstract target concepts. These concepts are generally difficult for students to understand. Being so, one of the important functions of analogies is to make the concepts difficult to understand become comprehensible. In this regard, most of the analogies used in physics textbooks were very reasonable in terms of the content of the target concept.

In terms of limitations of analogies, limitations were pointed out in 6% of the analogies used in physics textbooks, while there was no emphasis on the limitations in 94% of the analogies.

A similar study was carried out in Latvia (Jonane, 2015). Seven physics textbooks were examined regarding their usage of analogies. The analysis of analogies was directed toward their potential effectiveness for a deeper acquisition of science concepts and phenomena and for developing students' reasoning, meaning making, and transfer skills during teaching physics.

15.5.3 Social issues and 21st century skills in physics textbooks

Being among the most important social issues, sexism or "gender inequality" found a deserved place in *meta-curricular frameworks* for thinking about education and its objectives and praxis (Prazier and Sadker, 1973; and McCune and Matthews, 1975). No wonder some researchers interested in physics textbook analysis started to focus on this issue.

Taylor (1979) analyzed three physics textbooks used in England and found that there were frequent sexist biases as if physics were for boys and not for girls. The book "*The boy electrician*," published in 1920, was mentioned as an example that this biased vision has a long tradition.

Walford (1981) also found that physics learning materials, in their illustrations, questions and texts, presented physics as a male subject. He argued that an appropriate change in "image management" could encourage more girls to study physics.

Whiteley (1996), in his survey of seven physics textbooks in use in the Caribbean and Britain, gave further examples of gender imbalances. The more frequent depiction of males may have an adverse effect on the number of girls continuing their studies in physics.

Analysis of the gender issue in physics textbooks is still actual today (Rosa and Silva, 2020) and it is extended toward racial bias (Lawlor and Niiler, 2020).

15.5.4 21st century skills in physics textbooks

Google gives over six million hits for the search string "21st century skills." This means that these skills must be among the objectives of a *meta-curriculum framework*. Although there are different conceptualizations and definitions of these skills, almost all themes contain the skills related to critical thinking, collaboration, communication, and creativity. Many authors have written about how to teach and evaluate these skills (Rotherham, 2010; Care *et al.*, 2012; Kaufman, 2013; and Geisinger, 2016).

Critical and creative thinking are widely recognized as the most important 21st century skills for successful personal, professional, and social life (Trilling and Fadel, 2009; Larson and Miller, 2011; and Alismail and McGuire, 2015). Being an ideal context for learning and practicing these types of thinking, it is very surprising that there are only a few articles whose aim was to analyze how physics textbooks present the tasks of critical and creative thinking.

Klieger and Sherman (2015) carried out a study whose purpose was to understand whether and how physics textbooks (such as the Israeli high school book Newtonian Mechanics) enable the promotion and development of creative thinking. Findings indicated that they do not. This fact indicates that there is a need to raise physics teachers' awareness of the importance of creative thinking in learning materials.

Oktafianto *et al.* (2019) analyzed the presence of 21st century skills (creativity and innovation; critical thinking, problem solving and decision making; communication, and collaboration) in an Indonesian physics textbook for grade 10. They reported the following presence percentages: critical thinking skills and problem solving and decision making (61.86%), communication (15.81%), creativity and innovation (14.88%) and collaboration (7.44%). It is very strange that the authors didn't provide either verbal definitions or examples of physics textbook tasks (explanations, sample questions, students activities, exercises) that foster "creativity and innovation," "critical thinking, problem solving, and decision-making," "communication" and "collaboration."

15.6 ALIGNMENT OF PHYSICS TEXTBOOKS WITH INTRA-CURRICULAR PROCESSES

As commented before, physics textbooks are important causal factors in the processes of curriculum enactment and evaluation. For both of these *intra-curricular processes*, the quality of students' textbook readings determines the success of students' learning. Learning from physics textbooks is a complex interaction between the learner, text, and context variables. Being so, many researchers explored different aspects of that interaction.

In a review article, Alexander and Kulikowich (1994) presented a synthesis of the most important results:

a. Text processing in the domain of physics relies on readers' knowledge and interest, and on readers' ability to monitor or regulate their processing.
b. Certain textual features intended to assist readers in understanding and remembering physics content may work to the detriment of those very processes.
c. The inclusion of seductive details and the incorporation of analogies may misdirect readers' attention or may increase processing demands, particularly in those cases when readers' physics knowledge is low.
d. The questioning behaviors of teachers also impact the task of comprehending physics texts.
e. Within the context of the classroom, the information that teachers dispense or the materials they employ can significantly influence the process of learning from physics text.

15.6.1 Cognitive and textual aspects of textbook-based physics learning

Explorations of the process of the textbook-based physics learning were carried out in two ways. In the first way, researchers analyzed the cognitive adequacy of the physics textbook's tasks by supposing (not measuring!) the cognitive levels of the students (Prosser, 1979; Newton 1992; Khoja and Ventura, 1997).

In the second way, researchers included students in exploring the results of physics text readings.

Johnson (1979) selected and analyzed physics textbooks and examination papers. He found that three factors affect "readability." Two factors were related to the text ("legibility of print" and "complexity of words and sentences"). The students' related factor was "interest and motivation."

Dorothy Fagan (1997) explored the reasons why some students were unwilling to read physics-related materials. She found that two factors might work together to create unwillingness: "text difficulty" and "student's specific reading skills." She suggested that early detection of students with inadequate reading skills was important for giving them an extra teacher's help.

Adina Koch (2001) developed, applied and evaluated a metacognitive technique for improving student reading comprehension of physics texts. That technique required students to self-assess their reading comprehension and then rank their abilities and disabilities hierarchically. The technique was evaluated by comparing the performance of an experimental group on a reading-comprehension test of an experimental group with the that of a control group before and after the experimental manipulation. Both groups had the same reading-comprehension exercises, but only the experimental group received metacognitive tasks. Results showed that the posttest scores of the experimental group were significantly higher than those of the control group. Based on these results, Koch strongly recommended that the metacognitive technique should be developed and applied in teaching reading comprehension of physics texts as an effective self-monitoring device.

15.6.2 Potential relationship between textbooks and students' erroneous conceptual understanding

Acceptable level of textbook readability, commonly measured by the number of uncommon words and sentence lengths, doesn't necessary lead to successful physics learning. Inadequate conceptual and procedural treatments in textbooks might cause students to fail in gaining intended curriculum knowledge.

Dall'Alba and collaborators (1993) carried out important research, exploring the relationship between textbook treatments and students' understanding of acceleration. The sample included 60 final-year secondary (Year 12) and 30 first-year university students. They compared the ways in which acceleration is treated in physics textbooks with students' understanding of the same concept. It was found that some students' incomplete understandings of acceleration could be connected to misleading or inaccurate textbook treatments of the concept. Further limitations of some textbook treatments of acceleration were:

a. lack of attempts to make explicit relationships with other concepts;
b. failure to point out when it is appropriate to use particular definitions or that an alternative definition might be more appropriate in specific situations;
c. inclusion of operational definitions without conceptual explanations, and a focus on quantitative treatments while overlooking the development of qualitative understanding.

Not carrying out their own research on students' learning but overtaking the published results about students' alternative conceptions about the observer's eye in image formation, in a textbook study (Gürel and Eryilmaz, 2013), the optics content in 10 physics textbooks (nine American and one

Turkish) was analyzed to find out if the treatments related to students' alternative conception. It was found that analyzed textbooks ignored the role of the observer's eye or did not specifically emphasize the image formation or observation process. The authors suggested that textbooks should be reviewed by experts and the role of the observers' eye should be considered especially at the introductory optics for students to have a better understanding of the optical phenomena.

The same approach was taken by Guisasola *et al.* (2013) in the case of electromagnetic induction (EMI) and Faraday's Law. Before content analysis of these themes in the most used general physics textbook for scientists and engineers, they reviewed research-based literature on students' conceptual, epistemological, and methodological difficulties while learning the same topic. The last two difficulties were in their review focus because the authors believe that students have to use scientific techniques (questioning, giving hypothesis, designing experiments, …) to understand a theory.

Among the mentioned students' conceptual difficulties were, for example, the following:

> A significant number of students use explanations based on transmitting a "force" or "contact with the field."
>
> Most students do not distinguish between the empirical level (voltmeter and ammeter measurements) and the interpretative level that uses concepts such as fields and electromotive force (emf).
>
> Many students are not capable of recognizing EMI when there is no induced current.

The most important conclusion of this research was Only a minority of textbooks considered the students' prior knowledge and alternative conceptions relating to the concepts and laws of EMI.

The right step forward to treat the relationship between students' alternative conceptions and corresponding textbook treatments were done by Caleon and Subramaniam (2013) for the case of the propagation of periodic waves. They experimentally compared the conceptual-change efficacy of "traditional text" (derived from standard physics textbooks) and "refutational text," both featuring the particle-spring model. It was found that "refutational text" was more effective.

In addition, the authors found that a refutational video and depicting animation of the particle-spring model further reduce students' alternative conceptions about the topic when used to supplement either the refutational or the traditional text.

15.6.3 Digital physics textbooks in intra-curricular processes

Digital textbooks are change-makers in the world of education, from the processes of design, production, and marketing to the ways in which they are used by teachers in teaching and by the students in learning (Weisberg, 2011; and Joo *et al.*, 2017). Although textual parts are commonly shared with printed textbooks, digital textbooks have a big advantage: the possibility of including different forms of multimedia extensions that may influence students' performances in "implemented curriculum" and "learned curriculum."

Multimedia-based learning (MBL), in its simplest form, occurs when learners build their mental models for understanding by processing a combination of words and a sequence of static pictures. Learning can be further enhanced by video, computer simulation and animation (Mayer and Moreno, 2002). MBL has a cognition-based theoretical foundation that has been experimentally verified in many studies (Mayer, 2009). MBL consumers and designers can count on practical guidelines (Clark and Mayer, 2016). It is important to stress that in the Physics Education Research literature, an identical acronym MBL stands for Microcomputer-Based Laboratory (Thornton, 1987; and Thornton and Sokoloff, 1990).

For the case of multimedia-based physics learning, a few experimental results are worth mentioning:

Learning with online multimedia is improved if known students' physics misconceptions are treated by a refutation text (Muller *et al.*, 2008).

Innovative images of energy may help students construct key concepts (Ametller and Pintó, 2002).

Illustrations of air molecule motion and revised supporting text are better for students' learning of sound standing waves of air columns in pipe than traditional textbook text and illustrations (Zeng *et al.*, 2014).

Augmented reality technology also shows promising possibilities to enrich physics textbooks and improve learning (Bakri *et al.*, 2019).

When compared with textbook-based learning, new learning materials show better efficacy for both multimedia modules (Steltzer *et al.*, 2009) and electronic textbooks (Suyatna *et al.*, 2018).

Multimedia-based online prelectures can be more effective than textbook reading assignments (Sadaghiani, 2012). Before-class modules, usually 12–15 min in duration, introduce the key physics concepts through Flash animation with synchronous narration. The content of multimedia learning modules (MLM) is based on current research in multimedia learning. Research shows that the students who used the multimedia learning modules performed significantly better than students who did not view the MLMs, both on a test that was given immediately after viewing the MLMs and one given two weeks later.

A digital physics textbook makes possible the use of a Social Annotation Platform (SAP) for pre-class reading in a flipped introductory course (Miller *et al.*, 2018). The SAP platform allows students to discuss the reading online with their classmates and can be used as a research instrument to understand how students are reading before class.

It was found that, due to the use of SAP:

a. students spend an above average amount of time reading (compared to that reported in the literature);
b. most students complete their reading assignments before class; and
c. students employ active reading strategies and are involved in high-quality learning interactions outside class.

Exam performances of two cohorts of students, where the only difference between them is the use of the SAP, show that students do significantly better on exams when using the SAP.

Digital physics textbooks installed on personal computers make possible one-click connections with related, internet-based learning environments such as virtual laboratories, simulations, interactive videos, and intelligent tutors. These environments enhance students' active participation in intra-curricular processes (from enacted to learned curriculum).

15.7 CONCLUSIONS

This chapter presents basic (and, in a sense, problematic!) issues related to physics textbooks and curriculum alignment. In a very simplistic view, the alignment means only the relationship between "Intended Curriculum" and "Potentially Implemented Curriculum" as formulated by the textbooks.

It is quite clear that the analysis of that curriculum-textbook relationship is only possible in those educational systems where an official "Intended curricula" exists for physics.

The process of alignment analysis has two basic forms that are carried out by the different groups of persons with different purposes.

The first group are anonymous persons in the Ministry of Education who are in charge of verifying whether physics textbooks elaborated by interested publishers satisfy curricular aims and contents. Only those textbooks that pass that examination are approved for the use in schools. Although this process is very important, it is generally very secretive. Consequently, it was not possible to include some examples of that type of analysis in this chapter.

The second form of curriculum-textbook analysis is carried out by those researchers or curriculum specialists who focus their attention on the relationship between an *official physics curriculum* and *already approved physics textbooks*. They usually publish the results and conclusions of the resulting studies in research or pedagogical journals. An interesting situation occurs when curriculum specialists (who aren't employees of the Ministry of Education!) conclude that both national curriculum and approved physics textbooks should be improved (Mehmud and Iqbal, 2018).

A very big limitation of this chapter is the restriction to include only those curriculum - textbook alignment studies that were written and published in English. Due to this restriction, the number of articles that explicitly treat physics textbooks and curriculum alignment is surprisingly small.

One possible explanation of this fact is that in the USA, where a major part of Physics Education Research is carried out, an official National Physics Curriculum does not exist. In such a situation, as it was commented and documented in the chapter, the physics textbooks actually play the role of the "intended curriculum."

Nevertheless, as it was already argued in the chapter, researchers interested in textbook analysis have alternative opportunities to evaluate textbooks from broader perspectives. The first perspective is the alignment of textbooks with different *meta-curriculum frameworks.* The second perspective is related to the roles textbooks play in *intra-curriculum processes*.

As examples of the first perspective, the chapter reviewed the articles that posed research questions such as:

> Do physics textbooks treat the nature of science correctly?
>
> Do physics textbooks promote socially desired standards (for example, science literacy, gender, and race equity or 21st century skills)?

The second perspective was represented by the reviewed articles that were interested in issues like the following:

> Do physics textbooks have adequate readability and cognitive levels?

Especially, important were those articles that, starting from research results on textbook-based physics learning, suggested changes in the content and features of physics textbooks (as enrichment by multimedia and online resources).

In the future, more experimental studies should explicitly explore the role of textbooks as an "independent variable" that causally shapes, together with other possible factors, how physics is taught (Enacted Curriculum) and how and what students finally learn from their textbooks (Learned Curriculum). This last process became very important recently in that segment of the flipped classroom movement (Jensen *et al.*, 2018; and Awidi and Paynter, 2019) in which students are supposed to learn basic physics contents not by watching video-recorded lectures but by pre-class reading of printed or digital textbooks.

Additionally, it seems there aren't professional discussions and agreements about what should be an agenda of "physics textbook research" if it pretends to be a recognized part of Physics Education Research:

> What might be the most relevant issues of "textbook research" to support physics learning and teaching?
>
> From which theoretical frameworks and by which methodologies these "textbook research" issues should be treated?

A specific research field enters its mature phase if there is a "community" consensus about what are important problems and acceptable approaches to solve them, something close to the notion of "paradigm" (Kuhn, 1966). A "community" is formed when there are established national and international research groups with regular professional meetings (congresses, conferences, workshops, thematic nets, …), a peer-reviewed journal or a journal section, a thematic issue in a recognized journal, review articles, and at least one monograph or handbook with a synthesis of the most important results.

In the case of physics textbook research, such activities and academic products are not visible inside the community of Physics Education Research.

REFERENCES

Abd-El-Khalick, F. *et al.*, J. Res. Sci. Teach. **54**(1), 82–120 (2017).
Alexander, P. A. and Kulikowich, J. M., J. Res. Sci. Teach. **31**(9), 895–911 (1994).
Alismail, H. A. and McGuire, P., J. Educ. Pract. **6**(6), 150–154 (2015).
Ametller, J. and Pintó, R., Int. J. Sci. Edu. **24**(3), 285–312 (2002).
Ardwiyanti, D. *et al.*, J. Pendidikan Fis. Indones. **17**(1), 22–30 (2021).
Arriassecq, I. and Greca, I. M., Sci. Educ. **16**(1), 65–86 (2007).
Awidi, I. T. and Paynter, M., Comput. Educ. **128**, 269–283 (2019).
Bahçıvan, E. and Eraslan, F., Balkan Phys. Lett. **19**, 126–128 (2011).
Bakri, F. *et al.*, J. Penelitian Pengembangan Pendidikan Fis. **5**(2), 113–122 (2019).
Bancong, H. and Song, J., J. Pendidikan IPA Indones. **9**(3), 351–360 (2020).
Brookes, D. T. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020148 (2020).
Caleon, I. S. and Subramaniam, R., Res. Sci. Educ. **40**(3), 313–337 (2010).
Caleon, I. and Subramaniam, R., Phys. Educ. **48**(5), 657–663 (2013).
Care, E. *et al.*, *Assessment and Teaching of 21st Century Skills* (Springer, 2012).
Chukwunenye, J. N. *et al.*, AFRREV STECH **7**(1), 81–91 (2018).
Clark, R. C. and Mayer, R. E., *e-Learning and the Science of Instruction. Proven Guidelines for Consumers and Designers of Multimedia Learning*, 4th ed (Wiley & Sons, 2016).
Dall'Alba, G. *et al.*, J. Res. Sci. Teach. **30**(7), 621–635 (1993).
De Pro Bueno, A. *et al.*, Enseñanza Cien. Rev. Invest. Exper. Didácticas **26**, 193–210 (2008).
de Pro Chereguini, C. and de Pro Bueno, A., Rev. Eur. Sobre Enseñanza y Divulgación Cienc. **8**(2), 149–170 (2011).
Devetak, I. and Vogrinc, J., *Critical Analysis of Science Textbooks*, edited by M. S. Khine (Springer, 2013).
Docktor, J. L. and Mestre, J. P., Phys. Rev. Spec. Top.-Phys. Educ. Res. **10**(2), 020119 (2014).
Donahue, D. M., History Educ. Q. **33**(3), 321–352 (1993).
English, F. W., *Deciding What to Teach and Test: Developing, Aligning, and Auditing the Curriculum* (Corwin Press, 2000).
Epstein, L. C., *Thinking Physics is Gedanken Physics* (Insight Press, 1995).
Etkina, E. *et al.*, *College Physics: Explore and Apply*, 2nd ed (Pearson, 2019).
Fagan, D., Phys. Educ. **32**(6), 383–386 (1997).
Fan, L., ZDM, **45**(5), 765–777 (2013).
Fey, C. C. and Matthes, E., *The Palgrave Handbook of Textbook Studies*, edited by E. Fuchs and A. Bock (Palgrave Macmillan, 2018).
Forjan, M. and Sliško, J., Latin-Am. J. Phys. Educ. **8**(2), 241–247 (2014).
Forjan, M. and Sliško, J., Eur. J. Phys. Educ. **5**(3), 20–31 (2017).
Forjan, M. *et al.*, Latin-Am. J. Phys. Educ. **8**(3), 383–388 (2014).
French, M. *et al.*, College Teach. **63**, 171–177 (2015).
Geisinger, K. F., Appl. Meas. Educ. **29**(4), 245–249 (2016).
Gilbert, J. K. and Reiner, M., Int. J. Sci. Educ. **22**(3), 265–283 (2000).
Gim, J., Sci. Educ. **25**(5), 575–610 (2016).
Gürel, D. K. and Eryilmaz, A., Hacettepe Üniv. Eğitim Fakültesi Dergisi **28**(28-2), 234–245 (2013).
Guzsasola, J. *et al.*, Eur. J. Phys. **34**(4), 1015–1024 (2013).
Halliday, D. *et al.*, *Fundamentals of Physics. Extended Version*, 8th ed (John Wiley & Sons, 2008).
Hubisz, J. L., Phys. Teach. **39**(5), 304–309 (2001).
Jensen, J. L. *et al.*, J. Sci. Educ. Technol. **27**(6), 523–535 (2018).
Johnson, R. K., Sch. Sci. Rev. **60**(212), 562–568 (1979).
Jonāne, L., Pedagogika **119**(3), 116–125 (2015).
Joo, Y. J. *et al.*, Comput. Human Behav. **69**, 83–90 (2017).
Kaufman, K. J., Kappa Delta Pi Rec. **49**(2), 78–83 (2013).
Khoja, S. and Ventura, F., Mediterr. J. Educ. Stud. **2**(2), 119–129 (1997).
Klieger, A. and Sherman, G., Phys. Educ. **50**(3), 305–309 (2015).

Knight, B. A., *IARTEM 1991-2016. 25 Years Developing Textbook and Educational Media Research*, edited by J. Rodríguez Rodríguez *et al.* (IARTEM, 2019).

Koch, A., Sci. Educ. **85**(6), 758–768 (2001).

Kuhn, T. S., *The Structure of Scientific Revolutions* (The University of Chicago Press, 1966).

Larson, L. C. and Miller, T. N., Kappa Delta Pi Rec. **47**(3), 121–123 (2011).

Lawlor, T. M. and Niiler, T., Phys. Teach. **58**(5), 320–323 (2020).

Lee, S. and Lee, B., New Phys.: Sae Mulli **69**(4), 429–438 (2019).

Lehrman, R. L., Phys. Teach. **20**(8), 508–518 (1982).

Li, X. *et al.*, J. Baltic Sci. Educ. **17**(2), 229–238 (2018).

Mayer, R. E., *Multimedia Learning*, Second edition (Cambridge University Press, 2009).

Mayer, R. E. and Moreno, R., Educ. Psychol. Rev. **14**(1), 87–99 (2002).

McCune, S. D. and Matthews, M., J. Teach. Educ. **26**(4), 294–300 (1975).

McDermott, L. C. and Redish, E. F., Am. J. Phys. **67**(9), 755–767 (1999).

Mehmud, T. and Iqbal, M., Global Soc. Sci. Rev. **3**(1), 141–159 (2018).

Miller, K. *et al.*, Front. Educ. **3**(Article 8), 1–12 (2018).

Muller, D. A. *et al.*, J. Comput. Assist. Learn. **24**(2), 144–155 (2008).

National Research Council, *National Science Education Standards* (The National Academies Press, 1996).

National Research Council (NRC), *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (The National Academies Press, 2012).

Newton, D. P., J. Res. Read. **15**(2), 117–119 (1992).

Ogan-Bekiroglu, F., J. Sci. Teach. Educ. **18**(4), 599–628 (2007).

Ohanian, H. C., *Physics. Expanded*, 2nd edition (W.W. Norton, 1989).

Okoronka, A. U. and Adeoye, F. A., Int. J. Phys. Chem. Educ. **3**(2), 75–83 (2011).

Oktafianto, W. R. *et al.*, Phys. Commun. **3**(2), 72–77 (2019).

Park, D. Y. and Lavonen, J., *Critical Analysis of Science Textbooks* (Springer, 2013), pp. 219–238.

Park, W. *et al.*, Sci. Educ. **28**(9), 1055–1083 (2019).

Park, J. and Yoo, J., New Phys.: Sae Mulli **61**(9), 862–875 (2011).

Planinšič, G. *et al.*, Curriculum of physics for secondary school (2008), see http://portal.mss.edus.si/msswww/programi2008/programi/media/pdf/ucni_nacrti/UN_FIZIKA_strok_gimn.pdf

Polikoff, M. S. and Porter, A. C., Educ. Eval. Policy Anal. **36**, 399–416 (2014).

Polikoff, M. S. *et al.*, Educ. Meas.: Issues Pract. **34**(3), 10–17 (2015).

Prazier, N. and Sadker, M., *Sexism in School and Society* (Harper and Row Publishers, 1973).

Prosser, M., Sci. Educ. **63**(5), 677–683 (1979).

Raycroft, M. A. and Flynn, A. B., Chem. Educ. Res. Pract. **21**(4), 1110–1131 (2020).

Review Committee, Phys. Teach. **37**(5), 297–305 (1999a).

Review Committee, Phys. Teach. **37**(5), 306 (1999b).

Rosa, K. and Silva, M. R. G. D., Phys. Teach. **58**(9), 625–627 (2020).

Rotherham, A. J. and Willingham, D. T., Am. Educ. **17**(1), 17–20 (2010).

Rutherford, F. J. and Ahlgren, A., *Science for all Americans. Project 2061* (Oxford University Press, 1990).

Sadaghiani, H. R., Phys. Teach. **50**(5), 301–303 (2012).

Seitz, P., J. Canad. Assoc. Curric. Stud. (JCACS) **15**(1), 72–90 (2017).

Serway, R. A., *Physics for Scientists and Engineers*, 3rd ed. (Saunders College, 1990).

Squires, D. A., *Curriculum Alignment: Research-Based Strategies for Increasing Student Achievement* (Corwin Press, 2009).

Stelzer, T. *et al.*, Am. J. Phys. **77**(2), 184–190 (2009).

Suyatna, A. *et al.*, J. Pendidikan IPA Indones. **7**(4), 391–398 (2018).

Swartz, C., Phys. Teach. **37**(5), 283 (1999).

Taba, H., *Curriculum Development: Theory and Practice* (Brace & World, Harcourt, 1962).

Taylor, J., Phys. Educ. **14**(5), 277–280 (1979).

Tesfaye, C. L. and White, S., High school physics textbooks. Results from the 2008-09 Nationwide Survey of High School Physics Teachers. Focus On, September 2010, 1–9 (2010), see https://files.eric.ed.gov/fulltext/ED511375.pdf

Thacker, B. A., Rep. Prog. Phys. **66**(10), 1833–1864 (2003).

Thao-Do, T. P. and Yuenyong, C., J. Appl. Sci. Res. **9**(4), 2575–2584 (2013).

Thornton, R. K., Phys. Educ. **22**(4), 230–238 (1987).

Thornton, R. K. and Sokoloff, D. R., Am. J. Phys. **58**(9), 858–867 (1990).

Trilling, B. and Fadel, C., *21st Century Skills: Learning for Life in our Times* (Jossey-Bass, 2009).

Tyler, R. W., *Basic Principles of Curriculum and Instruction* (University of Chicago Press, 1949).

Valverde, G. A. *et al.*, *According to the Book: Using TIMSS to Investigate the Translation of Policy Into Practice Through the World of Textbooks* (Springer Science & Business Media, 2002).

Velentzas, A. and Halkia, K., Int. J. Sci. Educ. **40**(1), 90–108 (2018).

Walford, G., Phys. Educ. **16**(5), 261–265 (1981).

Webb, N. L., Natl. Inst. Sci. Educ. Brief **1**(2), 1–8 (1997).
Weisberg, M., Publ. Res. Q. **27**(2), 188–196 (2011).
Whiteley, P., Phys. Educ. **31**(3), 169–174 (1996).
Yager, R. E., J. Res. Sci. Teach. **20**(6), 577–588 (1983).
Yager, R. E., J. Res. Sci. Teach. **29**(8), 905–910 (1992).
Yener, D., Asia-Pac. Forum Sci. Learn. Teach. **13**(1), 1–17 (2012).
Young, H., *University Physics*, 8th ed. (Addison-Wesley, 1992).
Zeng, L. *et al.*, Phys. Rev. Spec. Top.-Phys. Educ. Res. **10**(2), 020110 (2014).
Zewdie, Z. M., Am. J. Educ. Res. **2**(1), 44–49 (2014).
Ziebell, N. and Clarke, D., Stud. Paedagog. **23**(2), 175–203 (2018).

CHAPTER

16 ANALYSIS OF PHYSICS TEXTBOOK CONTENT

Xiaomei Yan, Yuze He, Jingying Wang, Xiying Li, Xiaomei Ping, and Danhua Zhou

Yan, X., He, Y., Wang, J., Li, X., Ping, X., and Zhou, D., “Analysis of physics textbook content,” in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 16-1–16-34.

16.1 INTRODUCTION

In recent decades, physics education has been advanced by both the development of physics and science education. At secondary educational levels, many countries have issued new national curricula, including those for physics. To name just a few, the U.S. published the Next Generation Science Standards in 2013, the UK published a new “science programme of study” for key stage 4 in 2014, and Singapore published the O-Level Electronics syllabus and Science syllabus in 2021. In response to the new national curricula, new physics textbooks were published to reflect the change in educational goals and new understandings of physics. At the tertiary level, the contents of physics textbooks have also changed in response to the advances in physics, the decline of undergraduate physics enrolment and adapt to undergraduates’ diverse learning needs. Thus, this chapter focuses on the contents of physics textbooks, at both secondary and tertiary educational levels using a systematic review of the research literature.

Physics textbooks, as an important educational resource, contain abundant information and have a great influence on classroom teaching and learning. Among the growing numbers of textbook analyses and research (Vojíř and Rusek, 2019), people have begun to pay attention to the importance of content in textbooks. In order to unpack the multi-faced contents of physics textbooks, this study reviewed the contents of physics textbooks from conceptual, epistemic and social levels. These levels are related to the integrated goals of science education, including physics, proposed by Duschl (2008).

At the conceptual level, the great physics revolution of the 20th century required updating the content of textbooks (Varvoglis, 2014). Our knowledge of physics is rapidly evolving and changing the contents

https://doi.org/10.1063/9780735425514_016

The International Handbook of Physics Education Research: Special Topics, Mehmet Fatih Taşar and Paula R. L. Heron

of textbooks. However, since then, the content of physics textbooks at the secondary level, which describes "classical physics" has hardly changed for more than a century (Docktor and Mestre, 2014). In contrast, the content of textbooks in other scientific disciplines, such as biology and chemistry, is constantly changing (Sleeboom-Faulkner, 2008; Teo *et al.*, 2014; and Kummer-Hannoun and Roux-Goupille, 2015). There is a long debate on the consensus of what content of physics should be taught in schools.

Moreover, advances in scientific knowledge inspire philosophers of science to reflect the relationships between the amount of knowledge and understanding of knowledge (Park, 2017). Reflections on the understanding of science have led to increased attention to epistemology in science education. The objects of physics education have also been expanded, especially in metacognition and epistemology, such as scientific methods and skills and the nature of science (Yun, 2020). Epistemology has been considered by educators to be an important part of physics education and needs to be taught directly or indirectly in the classroom (Galili, 2019). For example, Lee and colleagues (2021) conducted a review of measurements on epistemologies in science education and emphasized the importance of epistemology in science education.

Furthermore, the rise of social cultural perspectives in science education considers science textbooks as 'cultural supportive tools' with significant cultural missions. Chisholm (2018) reviewed the studies on representations of race, class, and gender in textbooks. Especially with the equal education call, the equal representations of minority groups in textbooks were emphasized. For example, Jensen *et al.* (2021) examined the absence of disability references in Norwegian textbooks. Moreover, the influence of social constructivism also brings new understandings of concepts and knowledge construction in science, including physics (Kelly and Green, 1988). According to *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), research on the social dimensions of scientific knowledge can be traced back to philosophers in the 19th century, such as John Stuart Mill, Charles Sanders Peirce, and Karl Popper (Helen, 2019). There are two major camps: one acknowledges the influence of social settings on the scientific inquiry process, while the other regards sociality as a fundamental aspect of knowledge. These different views lead to the selections of contents and representations of contents of physics textbooks.

Representations of content are also explored in this study. How the contents are presented influences the students' sense making process. As for science textbooks, the balance between the readability and the scientific validity was a long debate. The word difficulty and sentence complexity have been examined by researchers using different tools in science textbooks at secondary level (i.e., Chiang-Soong and Yager, 1993; and Omebe, 2014) and at the primary level (Newport, 1965). Besides the text level of representations, more and more researchers have focused on the multiple representations in science education (Treagust, 2008) and found that they can help students organize information, improve conceptual comprehension and support the deep level of information process (such as Butcher, 2006; and Homer and Plass, 2010). Opfermann *et al.* (2017) specifically reviewed different models and theories on learning with multiple representations and outlined their importance for physics education. Salloum (2021) argued the importance of intertextuality in textbooks and its role in supporting students' diverse learning needs. There is also growing attention on multiple representations in textbooks with the development of multimedia and the emergence of e-textbooks.

In addition to what has been studied on the contents of physics textbooks, this chapter is also interested in how the contents were studied. Content analysis (CA) is a research methodology "for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use." Scholars usually use CA to describe or explain the process of meaning making from communication perspectives within textbooks. For example, CA was used to determine textbook compatibility with scientific approaches (Kusumaningrum and Indriyanti, 2017) or alignment to standards. This chapter discusses how the various topics were explored with CA in on physics textbooks from a methodology perspective.

In summary, this chapter reviewed the studies on the contents of physics textbooks and explored the following questions in particular:

1. What research topics are studied in the content analysis of physics textbooks? How have these topics evolved over the years?
2. What analysis methods are studied in the content analysis of physics textbooks? How have these methods changed over the years?

16.2 RESEARCH METHOD

16.2.1 Literature retrieval and screening

To get a complete picture of the research development, we searched extensively in three databases, namely Web of Science, Scopus, and ERIC, since each has its own set of rigorous review mechanisms to ensure the quality of the papers included. After several rounds of trial and discussion, we chose the following search strategies. The range included "title, abstract, keywords" with the search combinations of "physics" or "physical science" and "textbook*" or "teaching material*." The last search was conducted in April 2022, and the search results are shown in Table 16.1.

A total of 3463 papers were found. Then, duplicated articles were removed, and 1901 articles remained. Then, the abstracts of these papers were downloaded and scanned. This review only includes empirical articles written in English, excluding the letters, comments, book chapters, and non-English articles. Moreover, although schools in some countries teach integrated science, including physics, this review focuses on physics textbooks as used in independent disciplines and does not include general science textbooks. Thus, 462 articles remained. These articles were downloaded for further screening. The papers that did not actually analyze the contents of textbooks were excluded. For example, we found that some articles focused on the development of physics textbooks (Rambe *et al.*, 2019). Alternatively, physics textbooks were mentioned in some papers as the research background, and the

Table 16.1
Search results in different databases.

Database	Document (Publication) type	Language	Number of papers
Web of science	Articles	English	880
Scopus	Articles	English	1483
ERIC	Journal Articles	English	1100

main research was conducted with other focus, such as the one with students' interviews (Eijkelhof, 1996). Additionally, the studies that focused on certain topics mentioned by the textbooks but failed to actually analyze the textbooks were excluded. However, there may be relevant papers that were missed. Thus, the important papers referred to by the studied papers were added. After two rounds of screening, 138 articles were selected for our final review. The paper selection process is shown in Fig. 16.1. Throughout the screening process, two researchers read and selected all the papers independently. The items with different opinions were handed over to a third researcher for further discussion.

16.2.2 Coding system

Our study focused on the three levels of content and the representation of content. Each research paper was classified by the two researchers into one of the four categories that was the most appropriate. The consistency of the encoding is 0.94. When disagreements arose, the two researchers discussed with the third researcher to make a final decision.

16.2.2.1 Conceptual level of content

This category explores what is being presented in physics textbooks at the conceptual level, including the conception and learning context. Articles that explored the concepts, scientific terms, principles, laws, theories, and explanations as well as the contexts of the concepts (which are directly related to student learning) were included.

16.2.2.2 Epistemic level of content

This category covers the articles addressing broad topics beyond the conceptual level, including historical issues, philosophical issues, epistemological issues, and methodological issues. Notably, discussions on the nature of science and scientific methods were included in this category.

16.2.2.3 Cultural level of content

Critical studies on the content of physics textbooks are also interested in sociocultural issues, including multicultural and ethnic issues and gender issues. Studies addressing sex or racial bias and the representation of minority groups or women scientists in physics textbooks were included in this category.

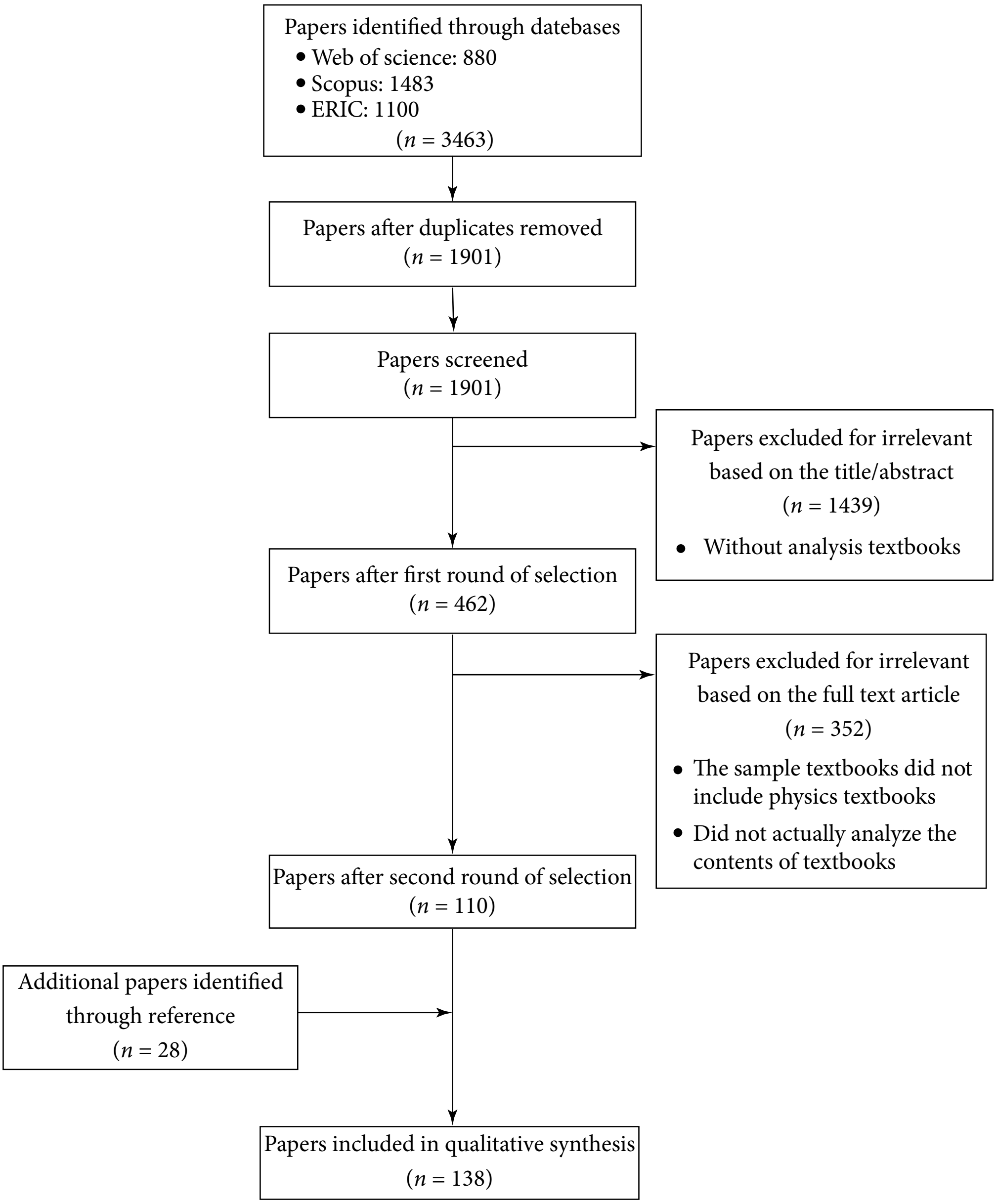


FIG. 16.1
The process of paper selection for the review.

16.2.2.4 Representational level of content

Studies on the representation of content in physics textbooks were included in this category. Studies in this category examined both verbal and visual representations in physics textbooks.

In the preliminary reading of the reviewed papers, we found that there are papers that addressed several themes. For example, Gauld (2004) analyzed old textbooks to discuss the different understandings of pendulums, which could be grouped into conceptual or epistemic levels. We discussed this case and coded it as epistemic level since the main discussions were about which theoretical perspectives were represented by the contents of the pendulum, rather than the scientific validity of the contents.

16.3 BASIC INFORMATION ABOUT THE REVIEWED LITERATURE

In this study, research papers focusing on the contents of physics textbooks indexed on Scopus, ERIC and Web of Science between 1940 and 2021 are presented. The number of publications and trends in each year (excluding 2022) is shown in Fig. 16.2.

Although papers were continuously published from the 1940s to the 1980s, 80% of the articles were published after the 2000s. The number of papers has increased rapidly since 2010, which aligns with the development trend of research papers in physics education (Yun, 2020).

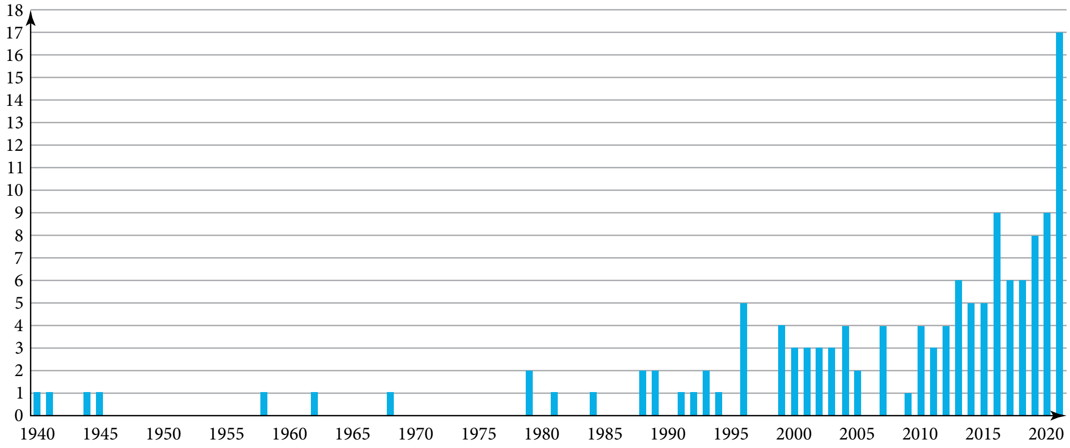


FIG. 16.2
Number of papers focused on physics textbooks CA in years.

16.3.1 Education level of papers

Studies on textbooks at the secondary education level (with 70 articles in total) outnumbered those at the tertiary level, as shown in Fig. 16.3. Thirteen articles focused on the transitions between educational levels, with two articles analyzing physics textbooks in both primary and secondary education levels and eleven articles analyzing physics textbooks in secondary and higher education levels at the same time.

Nevertheless, the studies at the tertiary level were published earlier than those at the secondary level, as shown in Fig. 16.4.

The interests of researchers at different educational levels varied in each period. Prior to 1999, there was little difference between researchers' interest in higher education and secondary education. However, in the decade 2000–2009, researchers were far more enthusiastic about higher education than about secondary education, and after 2010, this trend was reversed. A large number of studies at secondary levels were published in the last decade, which could be due to the educational reforms in many countries. There are a growing number of studies addressing the transitions between educational levels since 2000, which is align with the call of consistency science education curriculum and practices.

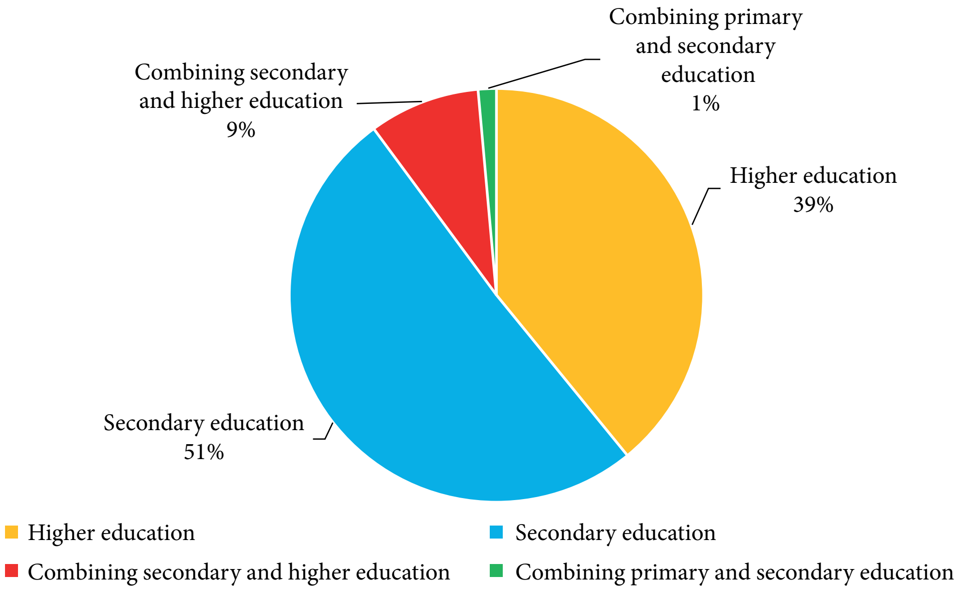


FIG. 16.3
Proportion of publications according to education level.

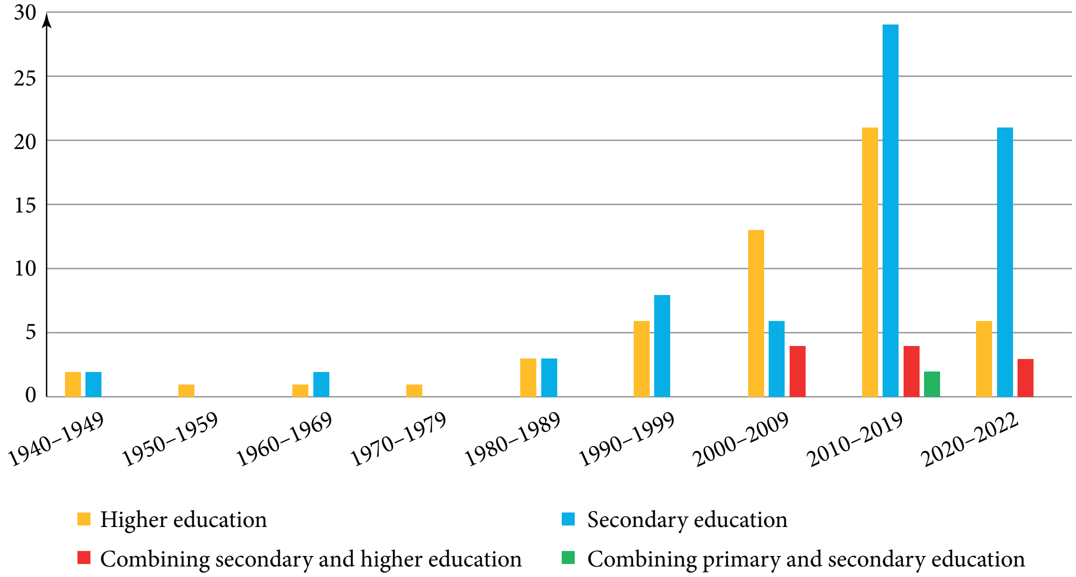


FIG. 16.4
The number of publications published in each decade according to the level of education.

16.3.2 Geographical distributions of papers

The reviewed studies also show the geographical diversity of the authors and research groups. We coded the paper's regional property according to the first author's affiliation, as shown in Fig. 16.5 and Table 16.2. Institutions from 36 countries contributed to this research area. The differences between countries are consistent with the overall paper output of researchers in different countries (Lin *et al.*, 2019). European institutions were the main contributors in this area, followed by North America and Asia. The high proportion of published papers from Europe and North America is similar to trends in science textbook research (Vojíř and Rusek, 2019). This could be attributed to the fact that there are a large number of English-speaking countries in these two regions, who are the main contributors to English-language journals. For the same reasons, American institutions contributed the most papers with 29 papers, followed by Australian institutions with 12 papers. However, the third most published country was Greece, with 11 papers, which is not an English-speaking country. Compared to the overall research output of Greek researchers, Greek authors seem to pay more attention to textbook research than other countries (Vojíř and Rusek, 2019). The proportion of Asia is almost the same as that of North America, which shows the emerging diversity in this area.

Furthermore, to investigate the international vision or global collaborations in this area, the regional property of the physics textbook was also coded according to the information in the main body or

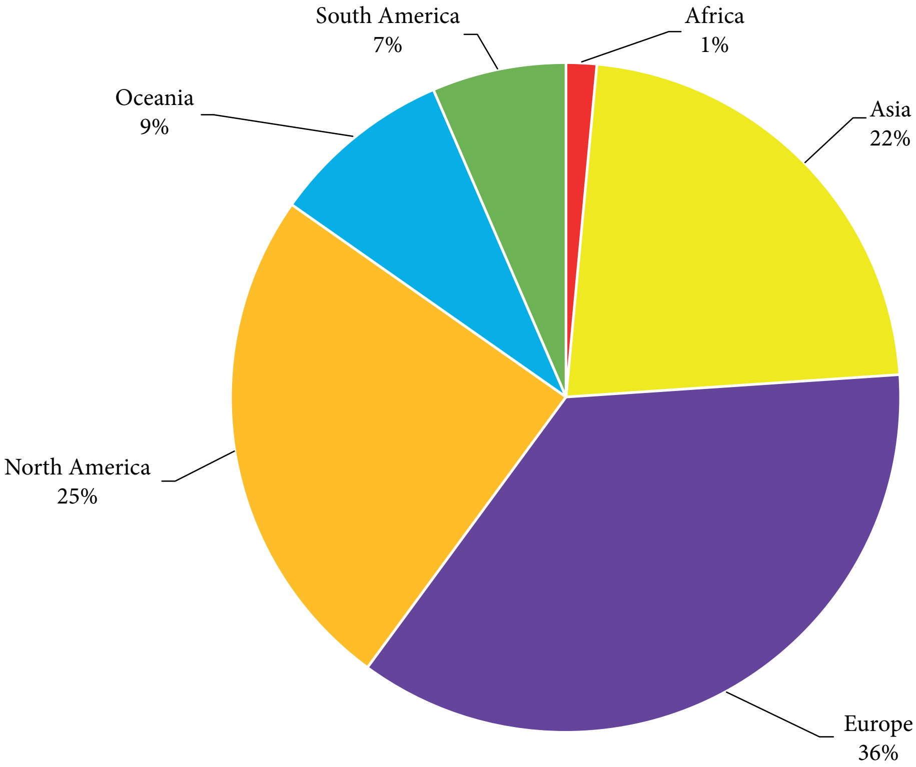


FIG. 16.5
The geographical distribution of the affiliation of the first author.

appendix of the paper. In all published articles, the content of physics textbooks in 36 countries was analyzed, since textbooks from different countries were compared in some articles. This study uses the number of articles to show the degree of researchers ‘attentions on physics textbooks in different countries. The results are shown in Fig. 16.6. Studies were predominantly on American textbooks, which could be partially due to the fact that historically, the U.S. published classical physics textbooks at the university level, such as *Physics Parts I and II* by Halliday and Resnick and *The Physics of every day Phenomena: A Conceptual Introduction to Physics II* by W.T. Griffith. It could also be related to the constraints of the technology; so English-language textbooks were used by international researchers. In particular, natural language processing (NLP) technology was developed for English but not for other Asian languages. For example, some Chinese researchers used NLP to study word occurrence and network modularity (Cui *et al.*, 2017) or knowledge structure (Cui *et al.*, 2014) with American physics textbooks. However, it is worth noting that in recent years, studies on Asian

Table 16.2
The country distribution of the affiliation of the first author.

Continent	Country	Number of papers	Continent	Country	Number of papers
Africa	South Africa	2	Europe	Belgium	1
South America	Brazil	3		Austria	1
	Argentina	4		Macedonia	1
	Venezuela	2		Ireland	1
Asia	Israel	2		Finland	2
	Pakistan	1		Czechia	1
	Lebanon	1		Slovakia	1
	Singapore	1		Latvian	2
	Korea	3		Portugal	3
	Türkiye	5		Sweden	1
	China	9		Norway	2
	Indonesia	8		France	3
	Iran	1		Germany	1
North America	Jamaica	1		Spain	5
	Canada	4		Russia	5
	United States	29		Greece	11
Oceania	Australia	12		United Kingdom	6
	Italy	2		Serbia	1

textbooks have grown rapidly. For example, in 2020 and 2021, 13 articles selected Asian physics textbooks for analysis, accounting for 50% of the total number of articles published. The textbooks studied are from China (Li *et al.*, 2020; Wang *et al.*, 2021; Wei *et al.*, 2022; Zhuang *et al.*, 2021; and Lin *et al.*, 2023), Korea (Kang, 2021), Pakistan (Hussain *et al.*, 2021), Singapore (Purysheva and Isaev, 2020) and Indonesia (McDonald and Abd-El-Khalick, 2017; Dewi *et al.*, 2020; Gumilar and Ismail, 2021; Sahriani *et al.*, 2021; and Yuni *et al.*, 2021).

The mismatch between the geographical distributions of the authors' affiliations and the textbooks studied shows that the authors did not solely focus on the textbooks of the country in which they were located. This could be partially attributed to the fact that the comparative study is often conducted in this field. For example, the Serbian author Mitrović (2020) compared the illustrations of physics textbooks from six countries, including Serbia, the United States, Russia, the United Kingdom, Bosnia and Herzegovina, and Belarus (Mitrović *et al.*, 2020).

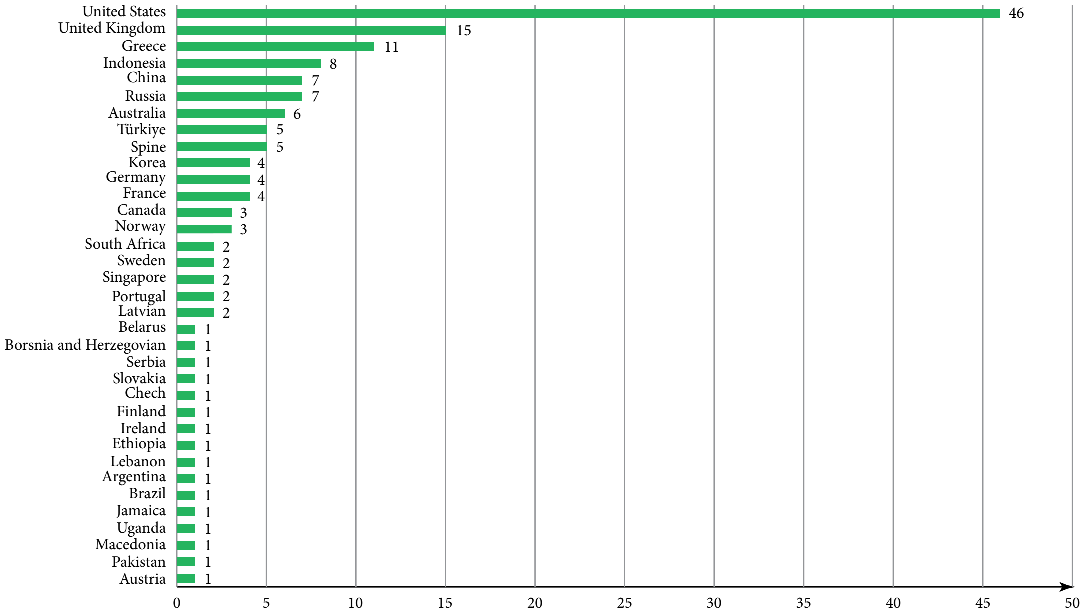


FIG. 16.6
Number of papers focused on physics textbooks in each country.

16.4 FINDINGS ON THE ANALYSIS OF PHYSICS TEXTBOOK CONTENT

16.4.1 Conceptual level of content

16.4.1.1 Basic information

There are 41 papers in total focused on what is being presented in physics textbooks at the conceptual level. Most papers studied textbooks at the tertiary level (24 out of 41). There are also studies exploring the transitions between education levels, such as five papers on textbooks from both tertiary and secondary levels and one paper on textbooks from both primary (science) and secondary levels. Moreover, most of the articles focus on a few concepts in the textbooks instead of the entire volume for analysis, which might be related to the in-depth elaboration on physics concepts.

The topic of the conceptual level of content has attracted researchers' attention since the 1940s, with at least one paper published every year since 2009. It is noticed that the researchers paid more attention to the physics textbooks of higher education in the past but shifted to secondary education since 2013. The early focus at the tertiary level could be related to the physics development at that time. Physics educators were concerned about the updates of new understandings in the textbooks (Docktor and Mestre, 2014). The interests at the secondary level might be related to the new curricula being published in the U.S. in 2012 and 2013, which brings new textbooks (Ndumanya *et al.*, 2021).

The authors in this topic are widely distributed, most of which are from Europe and North America (Fig. 16.7). Among them, authors from the United States were the main force on this topic. Ten of the 41 articles were written by authors from American institutions.

Except for a few articles that count the proportion of concepts in all physics textbooks (Weaver, 1945; and Başkan Takaoğlu, 2018), the other studies usually selected certain extracts with a focus on particular contents. Two main themes were identified under this category: the validity of the concepts and interdisciplinary concepts.

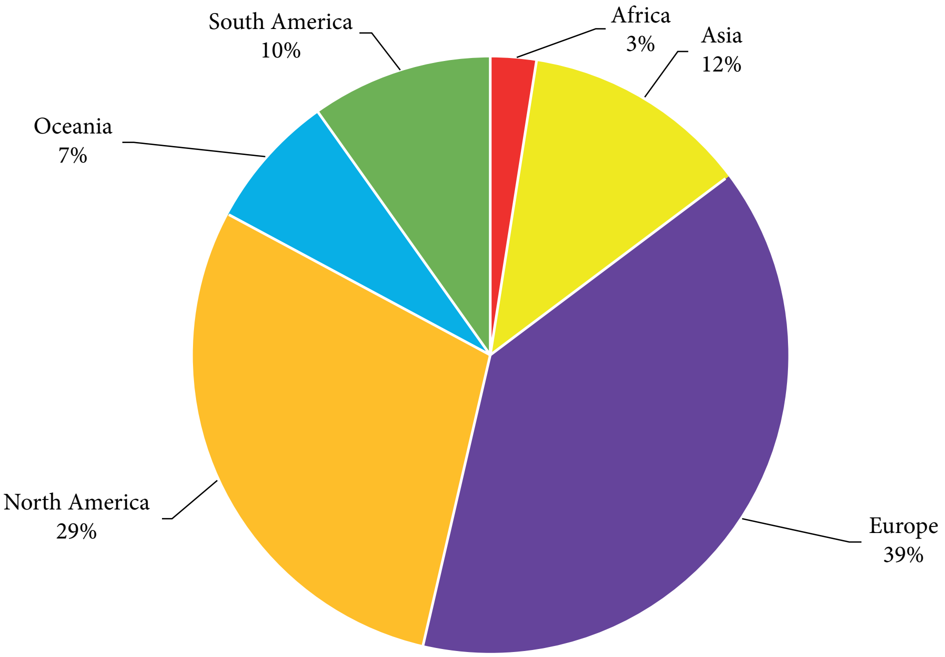


FIG. 16.7
The geographical distribution of the affiliation of the first author about learning conception.

16.4.1.2 Validity of concepts

Twenty-six of the 41 papers focus on the scientific validity of physical concepts in textbooks. Little (1940), Iona (1944), and Kenworthy (1941) started the research on the definition validity in the physics textbooks in the 1940s. For example, they analyzed the definitions of permeability μ and induction B (Little, 1940), the definitions of the potential difference (Kenworthy, 1941), and the definitions of mass, force and weight (Iona, 1944). All three articles discussed the consistency, completeness and correctness of the concept definition, which have been inherited in the follow-up research, although some studies only addressed one or two of them.

16.4.1.2.1 Inconsistent definitions

Inconsistencies in definitions across physics textbooks have been identified in many studies. For example, Kenworthy (1941) analyzed 22 college physics textbooks and found that three books ignored the definition of potential difference, while at least five different definitions appeared in the remaining 19 textbooks. The inconsistent definitions can still be found in recent physics textbooks. For example, the definition of "Pascal's principle" is still inconsistent in different textbooks (Anselmo *et al.*, 2020). To facilitate a more comprehensive analysis of conceptual consistency in physics textbooks, Stavrum *et al.* (2020) developed a conceptual framework to differentiate the definitions of "force" in physics textbooks: (1) Effect as property, (2) Force and motion, (3) Force is the cause of acceleration, (4) Force and Newton's second law, (5) Force and momentum, (6) Force and Newton's third law, (7) Force and work, and (8) Push-Pull. Based on this framework, Stavrum *et al.* (2020) identified the inconsistent "force" definitions in Norwegian middle school physics textbooks (Stavrum *et al.*, 2020).

Inconsistencies occur not only between books from different publishers but also across educational levels. Bächtold (2018) identified inconsistencies between primary and secondary textbooks in the definition of energy. In primary school, energy is mostly defined by Rankine's approach, while at the high school level, a definition from the conservation approach was introduced, at the expense of the definition from Rankine's approach (Bächtold, 2018).

Nevertheless, the inconsistency issues have been reduced in recent years. For example, Tural (2010) analyzed the definition of "weightlessness" in Turkish high school and university physics textbooks. Taibu (2015) selected 20 classic introductory physics textbooks to analyze the use of the term "weight." They came to a similar conclusion that the definition of "weight" in physics textbooks is consistent as the gravitational force of the earth on objects, mostly shown as $F = mg$ or $G = mg$.

16.4.1.2.2 Incomplete definitions

Incomplete definitions of concepts, were also identified in textbooks. Studies in the twentieth century found that the definitions of concepts in some textbooks were too simple. For example, Barnes (1958) analyzed 69 physics textbooks for the subject of impact and the coefficient of restitution in textbooks and found that much relevant information about concepts in these textbooks is missed. This incomplete issue persists many years later. For example, Kanderakis (2014) found that the concept of “work” was usually defined by its relationship with capacity transfer and kinetic energy change; it was described in a theoretical sense through formulas, while its implication was neglected.

16.4.1.2.3 Incorrect definitions

There are also errors in the definition of concepts identified in textbooks. Researchers usually compared a number of physics textbooks extensively for a certain concept. For example, Gearhart (1996) analyzed 27 introductory textbooks and found that only six presented the correct description of the law of energy equipartition. Gearhart (1996) further classified the errors into two types and inherited them by follow-up studies. One type of error is incorrect definition and formula. In this case, the textbook presents a completely wrong concept to students. For example, Sliško *ET al.* (2021) pointed out that typical physics textbooks attributed the cause of atmospheric pressure to the gravity of the air column above the surface, which confused atmospheric pressure with liquid pressure, ignoring the highly compressible nature of gases. The study then derived the atmospheric pressure formula through the formula to express that the interpretation of atmospheric pressure depending on the weight of the atmosphere is incompatible with theoretical calculations in physics textbooks (Sliško *et al.*, 2021). The other kind of error is incorrect context for the concept. Although the concept itself is correct, the textbook presented an inappropriate context for the concept, which may mislead students. The supplemental content in the textbooks, such as examples and the context/background information, also influences students’ meaning-making process. For example, Dall’Alba *et al.* (1993) identified the inaccurate descriptions of object motions in textbooks on the final year at the secondary level and introductory textbooks at the college level. Dall’Alba *et al.* (1993) found that textbooks stated that acceleration rather than force causes objects to fall, which makes it easy for students to mistake the relationship between physical concepts. Similarly, one piece of supplemental content corresponding to electromagnetic induction (EMI) and Faraday’s law based on the magnetic forces acting on the charge, but did not take into account the student’s knowledge and disconnected these concepts from scientific ideas (Guisasola *et al.*, 2013). Zajkov *et al.* (2017) claimed that the insufficient content of electromagnetic induction in secondary school textbooks encouraged students’ mechanical memory of the concept rather than inspired their deep understanding. Identifying these incorrect definitions help the researchers to offer focused improvement suggestions. Researchers usually offer a correct answer through formula derivation to the first type of errors (Ruddock, 2009; and Gezerlis; and Williams, 2021) and a better way of describing the context based on a literature review with contextual errors (Dall’Alba *et al.*, 1993; and Zajkov *et al.*, 2017). For example, the discussion of energy in the newly published physics textbooks extends to climate models (Brecha, 2021).

16.4.1.3 Interdisciplinary concepts

Among the 41 papers, six were concerned with the integration of knowledge of other disciplines, such as mathematics, into physics textbooks. This is a new area of research in recent physics textbooks as discussions on this topic first appeared in 2015 (Radtka, 2015). Mathematical tools are usually indispensable to construct explanations of physics, and mathematics is also one of the disciplines most closely related to physics (Başkan Takaoğlu, 2018). Through content analysis of physics textbooks from a mathematical perspective, it was found that common texts used for physics instruction tended to emphasize adding up pieces and procedural representations over the perimeter and area (Başkan Takaoğlu, 2018). However, this integration is actually difficult for students. When discussing the content of electromagnetism, clarifying the integral as a sum was a necessary step. It cannot be taken for granted that students can understand this mathematical method (Pina and Loverude, 2020). This line of study usually compared the important contents (such as calculus) between physics and mathematics textbooks.

16.4.2 Epistemic level of content

16.4.2.1 Basic information

The topic has attracted researchers' attention since 1979 (Prosser, 1979), with at least one paper published every year since 2010. There are 49 papers in total focused on what is being presented in physics textbooks at the epistemic level. The year 2021 was one of the most-followed years for articles on this topic, with 11 articles published, accounting for 65% of all articles reviewed that year. Most papers studied the textbooks at the secondary level (29 out of 49). There are also studies exploring the epistemic consistency across educational levels, such as one paper on textbooks from both tertiary and secondary levels and one paper on textbooks from both primary (science) and secondary levels. For the analysis unit, since the content of epistemology permeates all aspects of physics textbooks, nearly half of the research analyses the whole volume of textbooks (21 out of 49).

The authors on this topic are widely distributed, though most are from Asia and Europe (Fig. 16.8). Compared with other themes, Asian authors are more prominent. Among them, there are six articles each from China and Indonesia.

The researchers have explored diverse topics, including the history and philosophy of science (HPS), the epistemology of science, nature of science, and science practice. However, the boundaries and distinctions between each topic here are not clear-cut, since researchers in the science education community have not used the term consistently and some of the terms are still evolving (Berkovitz, 2017). We classified the papers based on the keywords they used.

16.4.2.2 Scientific literacy

The articles that analyzed the epistemic level of content used various keywords, including "scientific epistemology," "scientific literacy," "experiment," "scientific method" and "skills." Research interests in scientific literacy have a long history. The earliest framework for the quantitative analysis of scientific literacy content in textbooks was developed by Chiappetta (1991). The framework has four dimensions: (1) Knowledge of science, (2) Investigative Nature of Science, (3) Science as a Way of Thinking, and

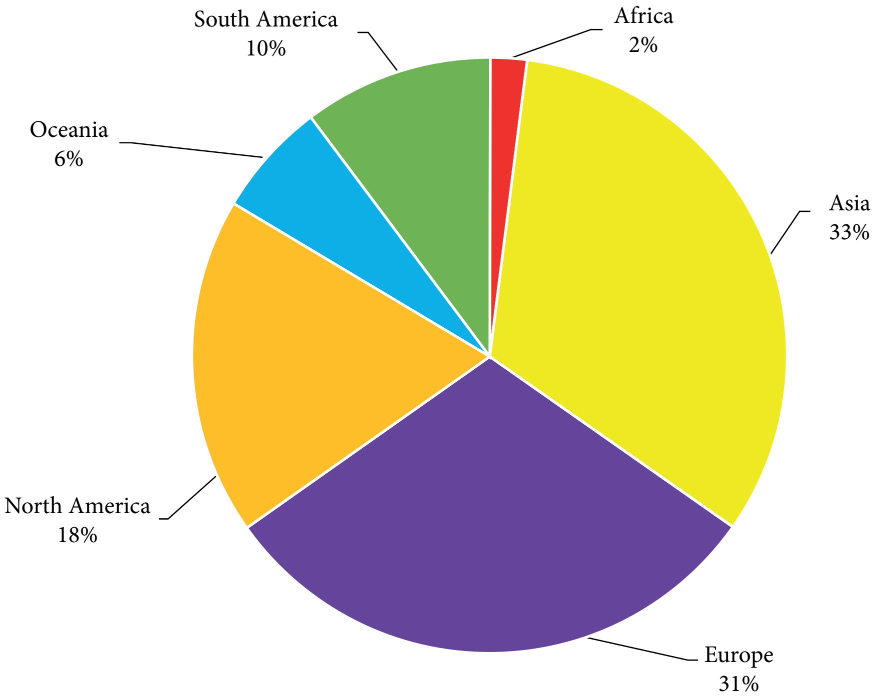


FIG. 16.8
The geographical distribution of the affiliation of the first author about History, philosophy, epistemology, and nature of science.

(4) Interaction Among Science, Technology and Society (Chiappetta *et al.*, 1991). In this sense, we regarded the studies of experimentation, scientific method and skills as part of scientific literacy.

Regarding the general analysis of scientific literacy, Chiappetta (1991) analyzed the physics textbooks used in American high schools, while Wilkinson (1999) used in Australian high schools and Sahriani *et al.* (2021) used in Indonesian high schools. They came to the same result: the proportion of "Knowledge of Science" dominated the textbooks (occupies 41.8%–61.6%) (Chiappetta *et al.*, 1991; Wilkinson, 1999; and Sahriani *et al.*, 2021). The proportion of "Science as a Way of Thinking" is neglected in textbooks, for example, none of the American physics textbooks (Chiappetta *et al.*, 1991) and only 6.7% in early Australian physics textbooks (before the 1990s) (Wilkinson, 1999). However, the representation of scientific literacy differs among countries and changes over time. For example, the least presented component of scientific literacy among Indonesian physics textbooks is "The Interaction Among Science, Technology and Society" (7.1%) (Sahriani *et al.*, 2021). In Australian physics textbooks, the representations of "Science as a Way of Thinking" increased, and "The Interaction Among Science, Technology and Society" (13.1%) became the least presented component of scientific literacy after the 1990s (Wilkinson, 1999).

In recent years, "Science as a Way of Thinking" has attracted research interest. For example, (Didiş Körhasan and Hıdır, 2019) analyzed analogies used in physics textbooks, which is regarded as a particular type of reasoning to build up physics knowledge.

Experiments are one of the important components of "Investigative Nature of Science." Experiments are of great significance to physics; in particular, some have claimed that classical physics is an experimental science (Vybiral, 2011). Although there is no clear definition of the word "experiment" in physics textbooks, experiments are still one of the indispensable contents in physics textbooks (Winston and Blais, 1996). However, the current research found that the experimental activities in physics textbooks are presented at a low cognitive level (Gumilar, 2021). Students must use the given methods following the prescribed procedures, as "follow- the-recipe, " with confirmative experiments. In addition to empirical experiments, thought experiments are explored in the reviewed papers. The results found that thought experiments are either neglected or insufficient presented in the current physics textbooks (Gilbert and Reiner, 2000; Velentzas *et al.*, 2007; and Bancong and Song, 2018). The results of the thought experiment are described as asserted rather than as anticipated (Gilbert and Reiner, 2000).

There are also studies examined 'The Interaction Among Science, Technology and Society (STS)" component. The contents of the textbooks revealed that the authors held different views on the interactions between technology and social development. Gardner (1999) discussed the meanings attached to the terms 'science' and 'technology' and outlines four views of their relationship: the idealist view, the demarcationist view, the materialist view, and the interactionist view. Five Canadian senior high school physics textbooks were analyzed and claimed to be mostly with an idealist view (Gardner, 1999). Researchers have also explored how society responds to the development of science and technology. For example, argued from a responsible citizenship perspective, Cottey (2010) examined 59 nuclear physics textbooks from 1950 to 2010 and found consistent features of unbalanced representations of nuclear power reactors and nuclear weapons in textbooks. Cottey (2010) argued that future citizens need to be better informed about the new developments in nuclear science and technology, including the threats, to make rational decisions in society. Therefore, regarding STS, the existing studies show that physics textbooks lack the contents of complex interactions among science, technology, and society.

16.4.2.3 History and philosophy of science (HPS)

Among the 49 articles, nine are devoted to HPS topics. Although the history of science and the philosophy of science are usually integrated, studies on textbooks have explored them separately. For example, there is a group of studies focused on history of physics. The earliest study in this group was conducted by Leite (2002), who examined the history of science with five Portuguese physics textbooks and proposed an analytical framework. This framework included eight dimensions: (1) Type and organization of the historical information, (2) Materials used, (3) Correctness and accuracy of the historical information, (4) Contexts to which the historical information is related, (5) Status of the historical content, (6) Learning activities dealing with the history of science, (7) Internal consistency of the book, and (8) Bibliography on the history of science. Twenty years later, this framework was adopted by Lin *et al.* (2023) to analyze the history of science in six Chinese high school physics textbooks. However, Lin *et al.* (2023) deleted "Correctness and accuracy" in the original framework and argued that there were multiple rounds of review by different departments to ensure the scientific correctness of history in the textbooks in China. They found that the history of science in the Chinese textbooks was too monotonous and rigid and lacked imagery and reality (Lin *et al.*, 2023). Unlike the quantitative approaches taken by Leite (2002) and Lin *et al.* (2023), two articles qualitatively discussed the related history of science in physics textbooks (Simon, 2016; and Montgomery and Kumar, 2021). Both Montgomery and Kumar (2021) and Simon (2016) advocated that physics textbooks should present historical information that helps students understand the process of scientific inquiry. Instead of analyzing the distribution of histories in the entire textbook, Persson (2018) focused on the history of the Planck blackbody radiation equation in physics textbooks (Persson, 2018). The above studies examining the integration of the history into physics textbooks did not involve discussions of the philosophy of science.

There are four studies that integrate the analysis of the philosophy of science and the history of science (Rodríguez and Niaz, 2004; Niaz *et al.*, 2010, 2013; and Klassen *et al.*, 2012). These studies developed frameworks to evaluate the integration level of history into three levels (satisfactory, mention, and no mention). They all concluded that the textbooks they examined often ignored the details of the history of science and philosophy of science (Rodríguez and Niaz, 2004; Niaz *et al.*, 2010, 2013; and Klassen *et al.*, 2012).

16.4.2.4 Nature of science (NOS)

Lederman (1992) defined NOS as the epistemology of science, which considers science as a way of knowing, or the values and beliefs inherent in the development and validation of scientific knowledge. Of the 49 articles, seven were specifically devoted to NOS. The presence and absence of NOS contents in physics textbooks is one of the research focuses. For example, Abd-El-Khalick *et al.* (2017) analyzed 34 popular American textbooks (16 biology and 18 physics) and found that, on average, less than 2.5% of the textbooks are dedicated to addressing NOS (Abd-El-Khalick *et al.*, 2017). Using the same frameworks, Li *et al.*(2020) and Zhuang *et al.* (2021) analyzed different versions of Chinese physics textbooks for junior high school and high school, respectively. The results showed that at least one dimension of the NOS was missing in each physics textbook, and there was a complete neglect of the consensus view of the nature of science, scientific theories, and scientific laws in some versions of physics textbooks (Li *et al.*, 2020; and Zhuang *et al.*, 2021).

Researchers are also interested in whether the NOS contents are presented explicitly or implicitly. Findings from most studies agreed that most the NOS in physics textbooks are implicit (Abd-El-Khalick *et al.*, 2017; Li *et al.*, 2020; and Zhuang *et al.*, 2021). The physics textbooks used in Indonesia represent all aspects of the NOS in a completely implicit manner (Ardwiyanti *et al.*, 2021). Moreover, NOS contents usually appear in either the preface or after word and rarely integrated with the main body in the textbook (Abd-El-Khalick *et al.*, 2017; and Park *et al.*, 2019). However, the nature of science is mentioned extensively in specific topics such as General Relativity Theory (GR) (Park *et al.*, 2019), which is inextricably linked to the fact that the topic of GR itself is full of philosophical thinking.

Studies on the NOS of physics textbooks also inherited the methodological challenges of NOS research in general science education. Philosophers, historians, sociologists of science, scientists and science educators have debated the exact definition of the NOS (Bell *et al.*, 1998). Guisasola *et al.* (2005) first analyzed the content of NOS in physics textbooks. Guisasola established an analytical framework with four dimensions of NOS to examine the contents of magnetic fields. The framework examined (1) the problem of the interpretation of magnetic interactions, (2) the construction of magnetic field theory, (3) the processes of unification, and (4) a critical view of the theory. However, the frameworks developed by Guisasola *et al.* (2005) have not been adopted by later researchers.

At present, some researchers have formed consensus views on the NOS. One of these is that it is difficult to reject the theory-driven nature of scientific observations (Abd-El-Khalick, 2012). Building on these consensus views, Abd-El-Khalick *et al.* (2008) developed an analytical framework for chemistry textbooks, and used in analyzing physics textbooks in 2017 (Abd-El-Khalick *et al.*, 2017). The framework includes 10 aspects: (1) Empirical, (2) Inferential, (3) Creative, (4) Theory-laden, (5) Tentative, (6) Myth of "The Scientific Method," (7) Scientific theories, (8) Scientific laws, (9) Social dimensions of science, and (10) Social and cultural embeddedness of science (Abd-El-Khalick *et al.*, 2017). In the follow-up studies, some researchers adapted this framework according to the characteristics of different textbooks and education levels (Li *et al.*, 2020; Ardwiyanti *et al.*, 2021; and Zhuang *et al.*, 2021). For example, ten NOS consensus views were used in the analysis of high school physics textbooks (Zhuang *et al.*, 2021), while seven of them were selected for the analysis of junior high school physics textbooks (Li *et al.*, 2020). Li and colleagues deleted three aspects of the NOS, including "theory-laden," "social dimensions of science," and "social and cultural embeddedness of science," which they claimed could not be understood by junior high school students (Li *et al.*, 2020). However, not all articles in recent years have used this framework. For example, Park *et al.* (2019) developed a particular NOS analysis framework for relativity theory. This framework contains a total of 11 dimensions: (1) aims and values, (2) methods, (3) scientific practices, (4) scientific knowledge, (5) social certification and dissemination, (6) scientific ethos, (7) social values, (8) professional activities, (9) social organizations and interactions, (10) financial systems, and (11) political power structures (Park *et al.*, 2019).

16.4.3 Cultural level of content

16.4.3.1 Basic information

There are 12 papers focused on representations of race, class, and gender in textbooks. Most of these papers were published in the past three years (6 out of 12). Most papers (10 out of 12) are focused at the secondary level. Ten out of 12 papers are from English-speaking countries. Also, there were no papers published by authors from Asian countries (Fig. 16.9).

The earliest studies in this area were conducted by Taylor (1979) and Walford (1981) on gender bias represented in the physics textbooks. Chisholm (2018) claimed that intense efforts to address the racism and sexism in textbooks are driven by the social movements in the 1960s and 1970s. The research interests continued until now. In the last 5 years, eight papers have been published. The rise of interest in gender issues again after 2010 could be related to the global focus on equal education and STEM education for underrepresented groups. In 2015, the United Nations proposed 17 Sustainable Development Goals (SDGs) in *Transforming Our World: the 2030 Agenda for Sustainable Development*, including "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all." This equal education call includes offering girls equal opportunities to learn physics. Additionally, with global interests in promoting STEM education, underrepresented female STEM workers have attracted attention.

The most commonly used method in this topic is through the analysis of pictures used in textbooks. Scientists in the textbooks were traditionally pictured as western, white, middle-class men, which was pointed out to be irrelevant to students from underrepresented groups, such as urban working classes,

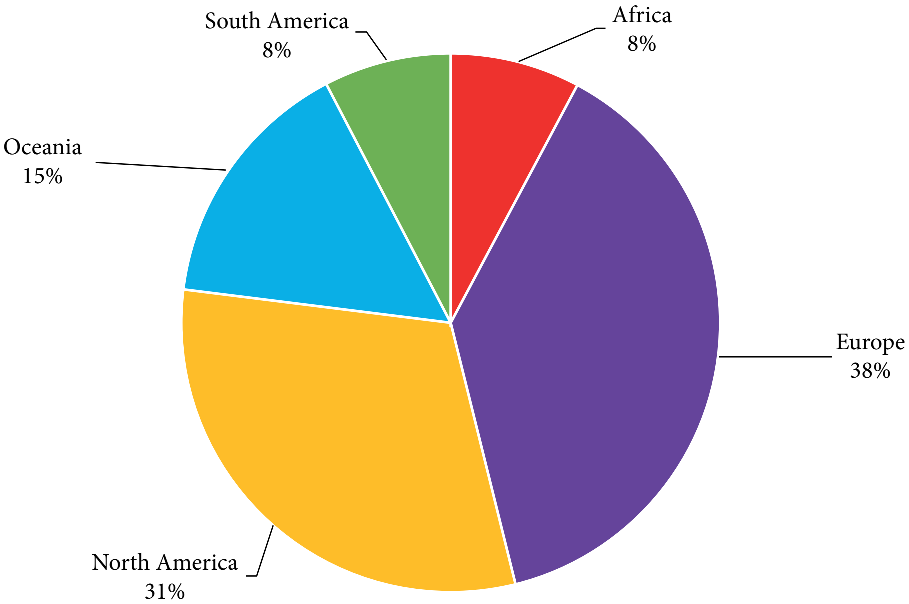


FIG. 16.9
The geographical distribution of the affiliation of the first author about cultural, social, and gender issues.

minority groups, or female students. Most studies used mixed methods, including quantifying the numbers of images, figures and illustrations and qualitative analysis of the meanings of the pictures in textbooks. For example, Lawlor and Niller (2020) coded racial representation by skin color. However, most of these studies focused on explicit representations. In contrast, Namatende-Sakwa (2019) took up feminist post-structuralism to conduct discourse analysis on both explicit and implicit representations and explored how gender is constructed in the five Ugandan physics textbooks at the secondary level. Namatende-Sakwa (2019) identified generic masculine epistemology under the mask of gender neutrality within physic textbooks.

16.4.3.2 Gender representations

Eleven out of 12 papers addressed the gender issues in the physics textbooks. In 1975, the Sex Discrimination Act became law. In 1979, Taylor (1979) explored the representations of females in the diagrams, photographs, and written words of three physics textbooks. Taylor found that females in scientific activities "were virtually nonexistent" (Taylor, 1979). In 1981, Walford (1981) examined the numbers of figures through illustrations, questions and texts in three physics textbooks published in 1979 and 1980. Walford found that the images projected by these newly published physics textbooks are still clearly biased towards boys. At the secondary level, Rosaand Silva (2020) analyzed three volumes of Brazilian high school physics textbooks and found that of the 154 images, only 33 were women and the rest were men (Rosa and Silva, 2020). Similarly, Keast (2017) examined the New South Wales physics textbooks and found that the content and imagery of male figures significantly outnumber females, which still present the field as a masculine domain.

Besides the study on representation bias, there is also an critical study on biased social roles reflected by the contents of the textbooks. Behrman (2017) identified the gender bias in the physics textbooks specially designed for female students in the 1910–1950s in the U.S. It is assumed that the students, who would have been future housewives, needed to be informed about the new technology and adapt to the new daily life with the technology development. Based on the analysis of the texts and images, Behrman (2017) claimed that the authors regarded female students (mainly white middle-class female students) as future users of new technology (most about household electric devices) at home.

16.4.3.3 Ethical representations

There is only one paper focused solely on the multi-cultural representations in the physics textbooks, together with the other two papers that mentioned ethical issues along with gender bias. Montgomery and Kumar (2021) examined three textbooks at the university level, which were found to have similar trends in repeating "tropes of extreme Eurocentrism" and minimal input from non-Western thinkers. Montgomery and Kumar (2021) argued that this lack of pluricultural content in textbooks for science undergraduates would profoundly influence contemporary scientists' fundamental and epistemological understanding of science and the process by which the relevant knowledge was actually created. This underrepresented issue is a historical challenge. For example, Lawlor and Niiler (2020) examined the distribution of images in 11 physics textbooks over a wide range of 60 years at the university level in the United States and found that ethnic or racial minorities are continuously underrepresented, along with women.

16.4.4 Representation of content

16.4.4.1 Basic information

There were 36 papers in total focused on how the contents are represented in the physics textbooks, including the linguistic features of language, the visual representation of the contents and the structure of contents. Twenty-five out of 36 papers analyzed the textbooks of middle schools, and the others analyzed textbooks at higher education levels. This area of research has attracted researchers' continuous interest since the 1960s, with rising interest in recent decades.

The authors in this area were distributed widely in the world, mostly from the European continent (Fig. 16.10), such as Greece, Russia, the Czech Republic, and Slovakia. There are also many papers from

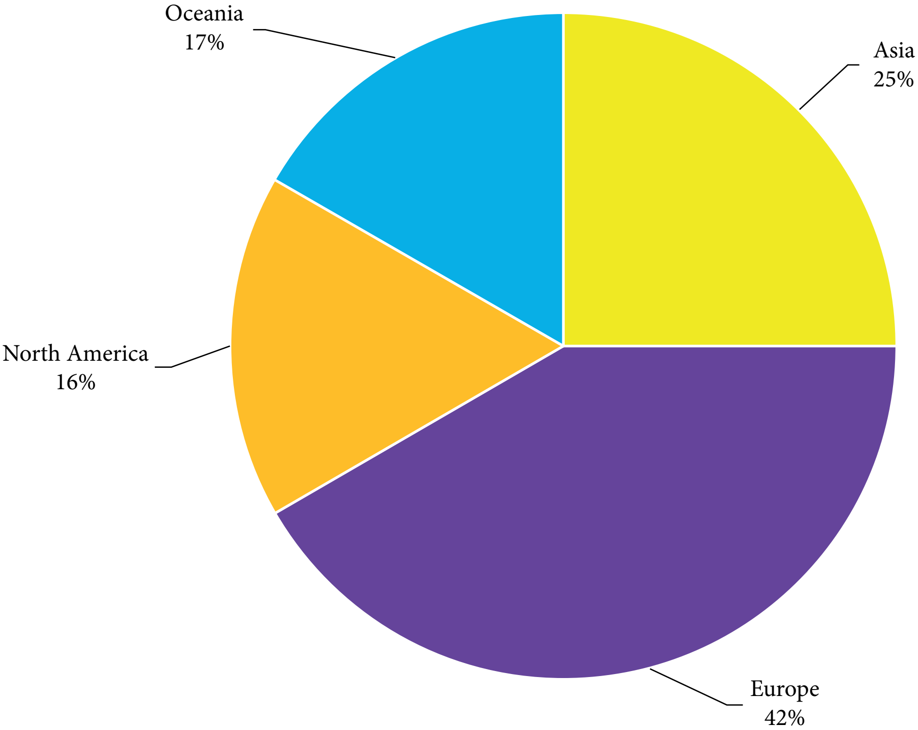


FIG. 16.10
The geographical distribution of the affiliation of the first author about representation.

Australia and the U.S., as well as China and Korea. In addition to researchers from science education and physics, this topic attracts linguist researchers such as Hussain *et al.* (2021) and Green (2019).

Comparison studies are often employed by researchers in this area. For example, researchers compare language features among different genres (Vuković Stamatović, 2020) or compare them to history textbooks (Ribeck and Borin, 2014). Moreover, with technology development, manual analysis has been replaced by automatic analysis (e.g., Hussain *et al.*, 2021), especially network analysis (e.g., Cui *et al.*, 2017; and Kralikova and Teleki, 2019) and natural language processing (e.g., Green, 2019) to reveal the linguistic and structural features of the contents. These new methods are believed to improve the efficiency and validity of research on language.

Various research topics were explored through the studies on the representations of the contents. For example, Elena (2016) explored the coherence between chemistry and physics to respond to the call for interdisciplinary studies in science education. However, one of the most interesting topics concerns the readability of the contents. It has been a challenge for authors and teachers to balance the scientific validity of the contents and readability of the texts.

16.4.4.2 Readability of contents

Researchers have investigated lasting endeavors to make sense of the readability of physics textbooks. As early as the 1960s, Marshall (1962) was interested in this topic and investigated the feasibility of the Flesch Reading Ease (FRE) formula in evaluating the readability of physics textbooks. FRE was also employed by Lanka *et al.*, (2013) in 2013 to examine the readability of Latvian Form 11 physics textbooks and found a mismatch with students' written language. However, Skorecova, Teleki, Lacsny & Zelenický (2016) proposed an easier method of "the probability distribution of words," compared to readability formulas, to explore the readability of two textbooks written in Slovakia and verified using eye-tracking techniques.

Some physics education researchers attributed the difficulties of understanding physics content to the abstractness and complexity of language. For example, Mayer (2016) assessed the complexity of the statements (including laws, principles, postulates) based on an abstraction scale. Moreover, some researchers are concerned about the validity of the scientific terms being used. For example, Leite (1999) explored the validity of terms at the secondary level, while Williams (1999) and Pina and Loverude (2020) examined the terms in *Introductory Physics* at the university level. Furthermore, beyond the word, terminology, and phrase levels, researchers have also explored how argumentation and explanations are built (Triantafillou *et al.*, 2016) and structures of passages (Ormiston-Smith, 1993; and Cui *et al.*, 2014).

Various linguistic aspects are explored by the researchers to investigate how students make sense of the texts. For example, Strube (1989) explored the language features of physics textbooks, including the prose structure, word choice, and literary characteristics. Hussain *et al.* (2021) explored the four-word Lexical bundles (LBs) such as "on the other hand," "at the same time," "on the surface of," and "at the end of" and their functional taxonomies employed in Pakistani Higher Secondary School Certificate (HSSC) level textbooks. From a sociolinguistic perspective, Dimopoulos *et al.* (2005) analyzed the classification, framing, and formality of texts and discussed the roles of readers (i.e., students) in relation to the body of knowledge.

16.4.4.3 Multi-representations of contents

In addition to the verbal representations, 4 (out of 36) papers discussed the visualizations of the contents. Three out of four papers were published in the last three years. This interest in the picture and illustration could be related to technology development and the growing literature on how multi-representation and visualization influence the meaning-making process. The studies extended the interests on students' understanding of verbal texts to visual representations. In particular, these researchers examined the illustrations of students' common misunderstandings in textbooks, with one on observing models (Mitrović *et al.*, 2020) and one with nuclear models (Hejnova and Králík, 2019). In contrast to their qualitative analysis, Kang (2021) conducted mixed methods of quantitative and qualitative analysis on the types, visual elements and expression of infographic presentations on waves. In addition to the focus on the conceptual understandings, Bungum (2013) was interested in how virtual representation influence students' attitudes toward physics. Bungum (2013) analyzed the "framing" of the pictures and discussed how "inviting" the pictures are and what kind of physics they "invite" the students to.

In addition to the research on visual and verbal representations, there are studies on multiple representations. For example, Salloum (2021) examined the levels of intertextuality, types of content, and quality of scaffolds in texts to explore how multiple representations and modalities can be better coordinated and synthesized for students' deep understanding.

16.5 CONTENT ANALYSIS OF KEY TECHNIQUES AND METHODS OF PHYSICS TEXTBOOKS

Various research methods were used to explore the features of the contents in physics textbooks. For example, early studies in the 1980s often used case studies to explore the validity of certain physics concepts. Strube (1989) explored the definitions and understandings of acceleration in textbooks and constructed categories of concept descriptions. These categories were drawn from the research data on the most distinctive features that differentiate one conception or way of understanding from another. The categories are presented in the form of a hierarchy, reflecting increasing levels of understanding. However, this type of case study with bottom-up data analysis has become rare in recent studies. Among the studies reviewed, content analysis is most often used.

Krippendorff (2019) defined content analysis (CA) as a research method and technology to make repeatable and effective data inferences. Bloor & Wood (2006) summarized the purpose of CA as describing the characteristics of the document's content by examining who says what, to whom, and with what effect. Although there are various understandings of content analysis and being used in different contexts in the research community, CA typically involves some guidelines for inference (based on existing theories, previous research, or experience) and strict procedural (i.e., coding) rules to move from unstructured text to answers to their research questions (White and Marsh, 2006). Along with the development of social science research, CA also experienced a change from quantitative analysis to qualitative analysis, covering broad research topics. This section focuses on how CA was used to inform the study of the contents of physics textbooks.

16.5.1 Theme-based quantitative CA

Theme-based quantitative content analysis or latent semantic analysis of explanation provides a reliable and clear way to describe what has been included in textbooks (Neuendorf and Skalski, 2009). In the articles reviewed, quantitative analysis based on the theme was broadly used to explore various topics, for example, gender equality (Whiteley, 1996), scientific literacy (Chiappetta *et al.*, 1991; Wilkinson, 1999; Alexiou and Skoumios, 2016; Rokhmah *et al.*, 2017; and Sahriani *et al.*, 2021), scientific inquiry (Li *et al.*, 2020), NOS (Guisasola *et al.*, 2005; Abd-El-Khalick *et al.*, 2017; Park *et al.*, 2019; Li *et al.*, 2020; and Ardwiyanti *et al.*, 2021), and intertextuality (Salloum, 2021).

Theme-based quantitative content analysis follows strict procedures. The most crucial step of quantitative content analysis is to construct a conceptual framework of textbook evaluation. According to the research purpose or education goals, an analysis framework (conceptual framework) is constructed. It was found that 36 reviewed articles adopted frameworks that were previously validated in other studies. These studies usually explored the topics of the nature of science (25 out of 36) and readability (8 out of 36). For example, following Lederman's Scientific Essence analysis framework, Zhuang *et al.* (2021) explored the NOS representations in five Chinese physics textbooks regarding the percentages, explicit levels, accuracy and completeness (simple, mixed, partially informed and informed), and overall consistency of different aspects of the NOS. Researchers employ analytical frameworks not limited to physics research. For example, studies draw on analytical frameworks originating from biology or chemistry textbooks. The most used frameworks were developed by Abd-El-Khalick *et al.* (2008), Thiele and Treagust (1994), Niaz (1998), and Chiappetta *et al.* (1991). These are all frameworks developed for the nature of science. These researchers argued that NOS is domain-general, which are widespread in textbooks across disciplines of science (Li *et al.*, 2020; Ardwiyanti *et al.*, 2021; and Zhuang *et al.*, 2021). However, researchers have emphasized on domain-specific features and advocated developing an analytical framework to fully capture the specific features of physics. Moreover, there were ten papers explored HPS topics with original frameworks, some of which were adapted by follow-up researchers. This continuity research effort can also be found on the topic of representation, while the FRE model was first proposed by Strube (1989) and inherited by recent researchers.

Moreover, 25 studies developed their own analytical frameworks. For example, to analyze how STEM is presented in Chinese science textbooks, Wang *et al.* (2021) developed an evaluation framework. High school physics, chemistry and biology textbooks from different publishers were explored using this framework. The results showed that there are differences in the total number, "location" and "degree of closeness" of concepts presented, and 'teaching objectives" in STEM disciplines. However, this kind of originally developed analytical framework requires a high theoretical competence of the researchers. Moreover, new frameworks (6 out of 25) are continuously advocated by linguistic researchers, which shows the robust and diverse research interests in this area.

16.5.2 Mixed methods with CA

In this study, there were 46 papers in total that used mixed methods of CA. Fewer and fewer studies employ strictly quantitative content analysis, while an increasing number of studies use both quantitative and qualitative methods to capture holistic or in-depth understandings of the contents in textbooks. For example, to explore the topics of representations or gender issues, mixed methods are often used to identify the percentage of target representation and qualitative discussions of what is being represented. Gumilar *et al.* (2022) compared representations of male and female in three Indonesian physics textbooks and identified gender stereotypes with qualitative example discussions.

Moreover, it is often collaborated with other methods. For example, Gumilar and Ismail (2021) used the content analysis method, i.e., the Inquiry Level Index (to categorize the students' compulsory laboratory activities of levels 0 to 3) and the Laboratory Assessment Inventory (which can be used to analyze students' activities and to determine which aspects dominate when conducting laboratory activities) as the main analytical frameworks for distinguishing levels of the activities in these textbooks. Complementary to these two content analysis of textbooks, the teachers' perspectives on delivering laboratory activities in class were obtained through semi-structured interviews.

16.5.3 ICT supported CA

Technology development also offers the opportunity to improve thve efficiency and effectiveness of content analysis methods. Traditionally, CA is conducted manually by researchers, which is time-consuming and involves researchers' subjective judgments. Addressing the methodological challenges of analyzing textbooks, automatic word analysis software such as CATPAC, WordStat, DLMAP, Concordance, General Inquirer, Atlastiand. For example, Hussain *et al.* (2021) explored four-word common Lexical bundles (LBs) and their functional taxonomies employed in Pakistani HSSC-level textbooks in chemistry and physics. A specialized corpus of these textbooks was built, which was run on Antconc software for the identification and extraction of the LBs in the corpus. The classification of the identified LBs was then carried out utilizing Biber functional taxonomies of LBs. The study generated a list of 102 four-word common core LBs, and all the LBs were analyzed functionally. The use of *Antconc* greatly improved the efficiency of analysis and validity. Moreover, the development of data science and network analysis offers further opportunities to explore the relationships and structural features of textbook contents. For example, Li and colleagues (Li *et al.*, 2020) used natural learning processing (lexicons, word embeddings, topic models) on 15 American history textbooks to explore the issues of gender, race, and ethnicity. It is expected that such computational toolkits will be more available and accessible to shed new light on textbook analysis, including physics textbooks. Some important issues are examined by the new technology. For example, Jiang and her colleagues (Jiang and Chen, 2019) developed a multidimensional network approach to explore the patterns of scientific knowledge diffusion, which could be used in future studies in textbooks.

Moreover, with the rapid development of technology in the past 20 years, new types of textbooks have emerged, including physics, such as various types of e-books, with multimedia information or even virtual reality (and augmented reality). The e-books claimed to be tailored to students' individual learning needs and to represent abstract physics knowledge in context, engaging students with an "authentic" context, etc. However, researchers have critiqued that technology-driven development neglects children's cognitive development rules and the needs of physics education. For example, the multimedia representations exceed the children's cognition load or neglect the opportunities for students to develop abstract and logical thinking on making sense of physics concepts. Therefore, there is growing research on using content analysis on ICT-enhanced textbooks, such as content analysis and evaluating e-textbooks (e.g., Peixinho and Vieira, 2015) and open textbooks (Hendricks *et al.*, 2017). It will be critical to conduct content analysis on exploring the features of multimedia representations in response to physics education challenges in particular.

16.6 CONCLUSIONS

This chapter reviewed the empirical research on the contents of physics textbooks and found a long history of research interests with contributions from research institutions around the world. Various hot topics in science education, such as the nature of science, argumentation, and interdisciplinary content, were identified in the review. The review was conducted to examine the conceptual, epistemic, social, and representational levels of concepts separately. Among the four levels, the epistemic level of content attracts the most studies (with 49 papers), while the cultural level of content attracts the fewest papers (with 12 papers). The focus on the conceptual level of content enjoyed the longest history, which dated back to the 1940s. The different educational levels also attracts different research interests. It is found that at higher education level, more focus is placed on conceptual level; while the secondary level is interested more on epistemic level.

Regarding the conceptual level of content, the findings suggest that more than 60% of papers are focused on tertiary-level education. Studies at the conceptual level are closely related to the development of physics. One of the main research topics concerns the scientific validity of the contents in physics textbooks. Early studies at the tertiary level concerned the inclusion of the new developments in the physics community. Studies on the scientific validity of concepts examined the consistency, completeness and correctness of definitions, explanations and contexts in the textbooks. The invalid concepts were persistently identified until recently in different countries. Inconsistent definitions of the same concept in different textbooks can cause confusion and learning difficulties for students (Dall'Alba *et al.*, 1993). However, Coelho (2010) argued that physics concepts are multifaceted, so it is reasonable for physics textbooks to present different concept definitions for different purposes. Constructing the definition of concepts that are consistent and multifaceted across educational levels is still a challenge for physics educators and textbook developers. Moreover, scientists' understanding of physics concepts has evolved along with the development of physics research. Therefore, the definition of concepts in physics textbooks is dynamic and complex, which requires the constant critical examinations. Furthermore, there is a historical challenge of integrating mathematics knowledge into physics textbooks. As early as the 1950s, Smith (1955) introduced a practice of integrated mathematics and physics courses. Feynman (1963) emphasized the essential roles of mathematics in the physics research process and justified the necessities of physics students to master mathematics. Monk (1999) suggested integrating algebra models in physics textbooks. However, the articles reviewed found there are inconsistencies between mathematics learnt in physics and mathematics textbooks or insufficient mathematics included in the physics textbooks. Similarly, Pospiech *et al.* (2019) found the same problem in their book on mathematics in physics education. Instead of regarding mathematics as the obstacle for students of learning physics, Pospiech *et al.* (2019) argued that mathematics could serve as an effective learning tool for students' better understanding of physics concepts. In addition, the articles that point out deficiencies in the concepts in physics textbooks demonstrate the critical reading of texts and offer opportunities to break the authority of textbooks. This practice can support teachers to use textbooks critically in class (Bansiong, 2019).

Regarding the epistemic level of content, the finding shows that most studies on this topic (29 out of 49) are at secondary level education. Along with the development of science education, the studies focused on research hot topics, such as NOS, HPS, and scientific literacy. There are abundant studies on scientific literacy-related contents, either studying the representations of different components of scientific literacy in the whole textbooks the particular parts of scientific literacy, such as the experiments or STS. The history contents are outnumbered in the studies of HPS, while some researchers have also suggested reviewing the contents from historical perspectives. For example, the editor of Der Karlsruher Physikkurs (KPK) physics textbooks (Herrmann, 2000) in Germany discussed how the different perspectives on the history of science influenced the structures of content. An effective way to bridge the gap between school science and what scientists actually do is to bring scientific concepts to science learning in a way that humanizes the protagonists and provides an appropriate context through the inclusion of an accurate history of science (Niaz *et al.*, 2010). Studies on NOS found that textbooks in different countries generally implicitly present some parts of NOS. The content analysis was used in different ways within this category. Some research explored how the key instances of the history of physics represented textbooks (such as the nature of light), while the others scanned the entire volumes of the textbooks for distribution features of the nature of science in the textbooks.

Regarding the cultural level of contents, there were only 12 papers on this topic focused on the stereotypes of scientists, with most examining gender bias. Gender bias was identified in the studies from 1979 to 2022 in the textbooks widely used in the U.S., Europe and Africa. However, there were no studies in Asia. Moreover, there were very few studies that addressed cultural diversity in the physics textbooks.

Regarding the representational level of content, the readability of the content is a critical and lasting research topic. Linguistic features, including the abstractness of words, scientific or popular terms, length of the words and sentences, and lexical bundles, are widely addressed. These findings on linguistic features contribute to the improvement of verbal representations of physics content. As a multi-representation collection, textbooks also support the study of discourse, explanation and argumentation. Research at the university level has explored the order and structure of topics, such as Majidi and Mantyla's (2011) study on organizing knowledge in physics textbooks. The same topic was analyzed by Tsaparlis (2014) for physics and chemistry textbooks. Moreover, the multi-representations were highlighted by researchers for physics and science education in general (Treagust and Tsui, 2013; and Opfermann *et al.*, 2017). The reviewed studies on non-textual explanation in the physics textbooks are focused on the illustrations of the experiments and the diagrams used in the process of scientific inquiry. Other science disciplines have explored the multi-representations in the textbooks more extensively from various theoretical perspectives. For example, Nyachwaya and Gillaspie (2016) studied the multi-representations in chemistry textbooks on the basis of cognitive load theory.

The findings also suggest that research on textbooks with content analysis methods is an important component of science education research and can be studied from various perspectives, such as the nature of science, scientific literacy, and the meaning-making process. On the other hand, the findings suggest that discipline-specific research on evaluating physics textbooks can promote the development of the techniques, process, methodological challenges and new technological tools for content analysis used in educational research in general.

16.7 LIMITATIONS AND IMPLICATIONS

There are limitations on the sampling process of this study that constrains the generalizability of the findings. First, this chapter focuses on the study of physics textbooks, and we excluded studies on science textbooks (integrated science textbooks including physics, such as Koliopoulos and Constantinou, 2005). However, many countries and regions teach the integrated science subject at the primary level and junior middle schools. Thus, our chapter focused on physics textbooks only at senior middle schools and higher education in many countries. Second, the studies excluded papers not written in English, which greatly decreased the international representativeness of this study. There are abundant and cutting edge studies in this area written in other languages. For example, Ibáñez-Ibáñez *et al.*, (2019) from Spain explored the view on NOS both in the theoretical content and the activities recommended in the textbooks. Cui, Zhu, Jung & Han (2016) tried a new algorithm in Korean to explore the bursty-word occurrence in physics textbooks. In the past two decades, 97 articles on physics textbook research have been included in the Chinese Social Science Citation Index (Zhang and

Yang, 2022). Third, the studies selected peer-reviewed journal articles and conference papers, while the volumes of books were neglected. These limitations of this review imply that future studies could be performed to expand the scope of the article reviewed, such as the relevant book chapters, degree theses and the studies involved in physics within general science textbooks. Reviews of non-English literature could also be conducted to draw a holistic map of international studies on the content of physics textbooks.

This review also indicates that future studies on the contents of physics textbooks could be conducted to address the challenge issues or researchers' different views. For example, more studies should be conducted on the conceptual and epistemic transitions of physics between educational levels to explore what should be taught at primary, secondary and tertiary levels. For example, Martins (2016) proposed reaching a consensus on what should be taught on the basis of an epistemology framework. More empirical research is needed to construct consistent and progressive physics textbooks. Moreover, abundant research and development in the field of physics education, especially different teaching theories (Alstein *et al.*, 2021), are not sufficiently reflected in physics textbooks. Physics textbooks at the secondary level are usually criticized for oversimplified content and often lack context and inspiration for imagination. At the same time, there are concerns regarding the limited space of physics textbooks, which prioritizes the presentation of concepts over the background information. The textbooks are regarded as a major source of information with accurate contents (Khine, 2013). The critiques argued that the context of concepts can be supplemented by teachers in the classroom. This debate raises the question of the role of textbooks in physics education, especially how to balance pedagogy and scientific information.

Moreover, collaboration among the educational, physics, and linguistics disciplines could be pursued to examine the contents from different perspectives. For example, although the studies on the conceptual level of content used formula derivation or literature verification to evaluate the concepts, most of the errors were manually identified by the researchers. The controversial understandings of the content are usually not responded to or verified by other researchers. Therefore, as discussed in Sec. V, emerging new technologies, such as various network analyses, could be employed to examine the features of textbook content and enhance the validity and efficiency of the studies on the conceptual level of content. However, the review of this study found that each discipline usually only published in its own journals or conferences. There were only a few studies reviewed in this chapter that borrowed linguistic tools from science educators to explore the features of content. For example, studies have explored how the linguistic features of texts influence students' interests in biology and physics textbooks (Mikk and Kukemelk, 2010). It is necessary to carry out more in-depth and comprehensive empirical research with other disciplines to integrate the scientific, pedagogical, and linguistic perspectives.

Furthermore, more comparisons could be made to identify the distinctive features of physics content in contrast to general science or other natural science disciplines such as chemistry and biology. Since the topics of HPS and NOS are extensively analyzed in other science textbooks (Abd-El-Khalick *et al.*, 2008), it makes sense to compare physics textbooks with other science textbooks. The existing research on HPS content in physics textbooks is very detailed, with many historical details discussed, but there still lacks a framework for the philosophy of science. Further studies are needed to explore how multi-representations could be integrated into physics textbooks and the representational competence in chemistry textbooks (Gurung *et al.*, 2022).

REFERENCES

Abd-El-Khalick, F., *Second International Handbook of Science Education*, edited by B. J. Fraser *et al.* (Springer, Netherlands, 2012), pp. 1041–1060. https://doi.org/10.1007/978-1-4020-9041-7_69.
Abd-El-Khalick, F. *et al.*, J. Res. Sci. Teach. **54**(1), 82–120 (2017).
Abd-El-Khalick, F. *et al.*, J. Res. Sci. Teach. **45**(7), 835–855 (2008).
Alexiou, N. and Skoumios, M., Int. J. Sci.: Math. Technol. Learn. **23**(1), 1–18 (2016).
Alstein, P. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(2), 023101 (2021).
Anselmo, D. H. A. L. *et al.*, Eur. J. Phys. **41**(6), 063001 (2020).
Ardwiyanti, D. *et al.*, J. Pendidikan Fisika Indonesia **17**(1), 22–30 (2021).
Bächtold, M., Res. Sci. Educ. **48**(2), 345–367 (2018).
Bancong, H. and Song, J., J. Pendidikan IPA Indonesia **7**(1), 25–33 (2018).
Bansiong, A. J., Cogent Educ. **6**(1), 1706395 (2019).
Barnes, G., Am. J. Phys. **26**(1), 9–12 (1958).
Başkan Takaoğlu, Z., Univ. J. Educ. Res. **6**(9), 1878–1886 (2018).
Behrman, J., Hist. Edu. **46**(2), 193–209 (2017).
Bell, R. L. *et al.*, J. Res. Sci. Teach. **35**: 1057–1061 (1998).
Berkovitz, J., Canad. J. Sci., Math. Technol. Educ. **17**(1), 37–45 (2017).
Bloor, M. and Wood, F., *Keywords in Qualitative Methods* (SAGE Publications Ltd., 2006).
Bungum, B., Phys. Educ. **48**, 648 (2013).
Butcher, K. R., J. Educ. Psychol. **98**(1), 182–197 (2006).
Chiang-Soong, B. and Yager, R. E., School Sci. Math. **93**, 24–27 (1993).
Chiappetta, E. L. *et al.*, J. Res. Sci. Teach. **28**(8), 713–725 (1991).
Chisholm, L., *The Palgrave Handbook of Textbook Studies*, edited by E. Fuchs and A. Bock (Palgrave Macmillan, New York, 2018).
Coelho, R. L., Sci. Educ. **19**(1), 91–113 (2010).
Cottey, A., Power Educ. **2**(2), 152–166 (2010).
Cui, X. *et al.*, Research on knowledge structure of physics textbook. 4 (2014).
CuiMarshall, J. S., Sci. Educ. **46**(4), 335–346 (1962).
Cui, X.-M. *et al.*, Phys. A: Statistical Mech. Appl. **487**, 103–110 (2017).
Cui, X.-M. *et al.*, New Phys.: Sae Mulli **66**(6), 762–767 (2016).
Dall'Alba, G. *et al.*, J. Res. Sci. Teach. **30**(7), 621–635 (1993).
Dewi, T. S. *et al.*, J. Phys.: Conf. Ser. **1521**(2), 022021 (2020).
Didiş Körhasan, N. and Hidir, M., Phys. Rev. Phys. Educ. Res. **15**, 010109 (2019).
Dimopoulos, K. *et al.*, Res. Sci. Educ. **35**, 173–195 (2005).
Docktor, J. L. and Mestre, J. P., Phys. Rev. Spec. Top.—Phys. Educ. Res. **10**(2), 020119 (2014).
Duschl, R., Rev. Res. Educ. **32**, 268–291 (2008).
Eijkelhof, H. M. C., Radiat. Prot. Dosim. **68**(3–4), 273–278 (1996).
Elena, I., J. Phys. Conf. Ser. **738**(1), 012105 (2016).
Feynman, R. P., *The Feynman Lectures on Physics* (Addison-Wesley, 1963), ISBN 978-0201020106.
Galili, I., Sci. Educ. **28**(3–5), 503–537 (2019).
Gardner, P. L., Int. J. Sci. Educ. **21**(3), 329–347 (1999).
Gauld, C., Sci. Educ. **13**, 321–332 (2004).
Gearhart, C. A., Am. J. Phys. **64**(8), 995–1000 (1996).
Gezerlis, A. and Williams, M., Am. J. Phys. **89**(1), 51–60 (2021).
Gilbert, J. K. and Reiner, M., Int. J. Sci. Educ. **22**(3), 265–283 (2000).
Griffith, W. T., *The Physics of Everyday Phenomena: A Conceptual Introduction To Physics*, second ed. (WCB/McGraw-Hill, Boston, Mass, 1998).
Green, C., Ling. Edu. **53**, 100748 (2019).
Guisasola, J. *et al.*, Sci. Educ. **14**(3–5), 321–328 (2005).
Guisasola, J. *et al.*, Eur. J. Phys. **34**(4), 1015–1024 (2013).
Gumilar, S. *et al.*, Int. J. Sci. Educ. **44**(3), 416–433 (2022).

Gumilar, S. and Ismail, A., Res. Sci. Technol. Educ. (published online 2021).
Gurung, E. *et al.*, J. Chem. Educ. **99**(5), 2044–2054 (2022).
Hejnova, E. and Králík, J., AIP Conf. Proc. **2152**, 030006 (2019).
Helen, L., *The Social Dimensions of Scientific Knowledge. The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, Summer 2019 ed. (Metaphysics Research Lab, Stanford University, 2019). See https://plato.stanford.edu/archives/sum2019/entries/scientific-knowledge-social/.
Hendricks, C. *et al.*, Int. Rev. Res. Open Distance Learn. **18**(4), 78–99 (2017).
Herrmann, F., Eur. J. Phys. **21**(1), 49 (2000).
Herrmann, F., Eur. J. Phys. **21**, 49 (2000).
Homer, B. D. and Plass, J. L., Inst. Sci. **38**, 259–276 (2010).
Hussain, G. *et al.*, GEMA Online **21**(1), 221–238 (2021).
Ibáñez-Ibáñez, M. *et al.*, Rev. Invest. Experiencias Didácticas **37**, 49 (2019).
Iona, M., Am. J. Phys. **12**(6), 368–369 (1944).
Jensen, M. S. *et al.*, Int. J. Inclusive Educ. (published online 2021).
Jiang, S. and Chen, H., Scientometrics **121**, 1599–1617 (2019).
Kanderakis, N., Sci. Educ. **23**(6), 1293–1308 (2014).
Kang, K., New Phys.: Sae Mulli **71**(11), 921–928 (2021).
Keast, V. J. J., History Educ. **46**(2), 193–209 (2017).
Kelly, G. and Green, J., *Perspectives on Conceptual Change: Multiple Ways to Understand Knowing and Learning in a Complex World*, edited by B. Guzzetti and C. Hynd (Lawrence Erlbaum Associates, London, 1988).
Kenworthy, R. W., Am. J. Phys. **9**(6), 380–381 (1941).
Khine, M. S., *Critical Analysis of Science Textbooks*, edited by M. Khine (Springer, Dordrecht, 2013).
Klassen, S. *et al.*, Sci. Educ. **21**(5), 729–743 (2012).
Koliopoulos, D. and Constantinou, C., *The Pendulum*, edited by M. R. Matthews *et al.* (Springer, Dordrecht, 2005).
Kralikova, P. and Teleki, A. (2019). New Trends and Issues Proc. Human. Soc. Sci. **6**(1), 30–37.
Krippendorff, K., *Content Analysis* (SAGE Publications, Ltd., 2019).
Kummer-Hannoun, P. and Roux-Goupille, C., Sci. Textbooks **7**(3), 29 (2015).
Kusumaningrum, I. *et al.*, J. Phys.: Conf. Ser. **895**, 012042 (2017).
Lanka, M. *et al.*, Proceedings of the International Scientific Conference (2013), Volume 6, 227–234.
Lawlor, T. M. and Niiler, T., Phys. Teach. **58**(5), 320–323 (2020).
Lederman, N. G., J. Res. Sci. Teach. **29**, 331–359 (1992).
Lee, S. W.-Y. *et al.*, Sci. Educ. **105**, 880–907 (2021).
Leite, L., Sci. Educ. **11**, 333–359 (2002).
Li, X. *et al.*, Res. Sci. Educ. **50**(3), 833–844 (2020).
Lin, T.-J. *et al.*, Int. J. Sci. Educ. **41**(3), 367–387 (2019).
Lin, T.-C. *et al.*, Int. J. Sci. Educ. **36**(8), 1346–1372 (2014).
Lin, L. *et al.*, Sci. Educ. **32**(1), 101–121 (2023).
Little, E. M., Am. J. Phys. **8**(2), 129–131 (1940).
Majidi, S. and Mäntylä, T., J. Baltic Sci. Educ. **10**(4), 15 (2011).
Marshall, J. S., Comprehension and alleged readability of high school physics textbooks (1962).
Martins, A. F. P., Res. Sci. Educ. **46**, 511–524 (2016).
Mayer, R. V., Conference Name 9th Annual International Conference of Education, Research and Innovation, Spain (IATED, 2016), pp. 14–16.
McDonald, C. V. and Abd-El-Khalick, F., *Representations of Nature of Science in School Science Textbooks: A Global Perspective*, 1st ed. (Routledge, 2017).
Mikk, J. and Kukemelk, H., Trames **14**, 54–70 (2010).
Mitrović, M. M. *et al.*, Am. J. Phys. **88**(2), 141–147 (2020).
Monk, M., Phys. Educ. **29**(4), 209 (1999).
Montgomery, S. L. and Kumar, A., KNOW: A Journal on the Formation of Knowledge **5**(2), 197–211 (2021).
Namatende-Sakwa, L., Gender Educ. **31**(3), 362–376 (2019).
Ndumanya, E. *et al.*, Can. J. Sci.: Math. Technol. Educ. **21**(3), 539–552 (2021).
Neuendorf, K. A. and Skalski, P. D., *Measuring Identity: A Guide for Social Scientists*, edited by A. I. Johnston *et al.* (Cambridge University Press, 2009), pp. 203–236.
Newport, J. F., Elemenary School J. **66**, 40–43 (1965).
Niaz, M., Sci. Educ. **82**(5), 527–552 (1998).
Niaz, M. *et al.*, Sci. Educ. **94**(5), 903–931 (2010).
Niaz, M. *et al.*, Phys. Educ. **48**(1), 57–64 (2013).
Nyachwaya, J. M. and Gillaspie, M., Chem. Educ. Res. Practice **17**(17), 58–71 (2016).
Omebe, C. A., Int. J. Sci. Eng. Res. **5**(12), 1059–1062 (2014).
Opfermann, M. *et al.*, *Multiple Representations in Physics Education. Models and Modeling in Science Education*, edited by D. Treagust *et al.* (Springer, Cham, 2017), Vol. 10.
Ormiston-Smith, H., Res. Sci. Educ. **23**, 222–227 (1993).

Park, S., J. Gen. Philos. Sci**48**, 569–579 (2017).
Park, W. *et al.*, Sci. Educ. **28**(9–10), 1055–1083 (2019).
Peixinho, J. P. and Vieira, R. M., *Inted2015: 9th International Technology, Education and Development Conference*, edited by L. G. Chova *et al.* (Iated-Int Assoc Technology Education & Development, 2015), pp. 3776–3783.
Persson, J. R., Am. J. Phys. **86**(12), 887–892 (2018).
Pina, A. and Loverude, M. E., *2019 Physics Education Research Conference Proceedings, Provo, UT* (AAPT, 2020).
Pospiech, G. *et al.*, *Mathematics in Physics Education* (Springer Nature, Cham, 2019).
Prosser, M., Sci. Educ. **63**(5), 677–683 (1979).
Purysheva, N. S. and Isaev, D. A., J. Phys.: Conf. Ser. **1691**(1), 012052 (2020).
Radtka, C., Sci. Educ. **24**(5–6), 725–748 (2015).
Rambe, A. *et al.*, J. Pendidikan Fis. Indones. **15**(2), 70–79 (2019).
Resnick, R. and Halliday, D., *Physics, Parts I and II* (Wiley, 1978).
Rosa, K. and Silva, M. R. G. D. Phys. Teach. **58**(9), 625–627 (2020).
Ribeck, J. and Borin, L., *Human Language Technology Challenges for Computer Science and Linguistics. LTC 2011. Lecture Notes in Computer Science*, edited by Z. Vetulani, and J. Mariani (Springer, Cham, 2014).
Rodríguez, M. A. and Niaz, M., J. Sci. Educ. Technol. **13**(3), 409–424 (2004).
Rokhmah, A. *et al.*, J. Pendidikan Fis. Indones. **13**(1), 19–24 (2017).
Ruddock, I. S., Eur. J. Phys. **30**(2), 303–309 (2009).
Sahriani, S. *et al.*, J. Phys.: Conf. Ser. **2098**(1), 012005 (2021).
Salloum, S., Int. J. Sci. Educ. **43**(17), 2814–2842 (2021).
Simon, J., Histor. Stud. Nat. Sci. **46**(3), 392–427 (2016).
Sleeboom-Faulkner, M., BioSocieties **3**(1), 21–36 (2008).
Sliško, J. *et al.*, Phys. Teach. **59**(6), 470 (2021).
Skorecova *et al.*, Phys. Educ. **51**, 065009 (2016).
Smith, M., Math. Teach. **48**, 535–537 (1955).
Stavrum, L. R. *et al.*, Nordic Stud Sci. Educ. **16**(2), 183–198 (2020).
Strube, P., Int. J. Sci. Educ. **11**(2), 195–202 (1989).
Taibu, R. *et al.*, Phys. Rev. Spec. Top. – Phys. Educ. Res. **11**(1), 010117 (2015).
Taylor, J., Phys. Educ. **14**, 277 (1979).
Teo, T. W. *et al.*, Chem. Educ. Res. Pract. **15**(4), 470–487 (2014).
Thiele, R. B. and Treagust, D. F., Instruct. Sci. **22**, 61–74 (1994).
Treagust, D., *Science Education at the Nexus of Theory and Practice*, edited by Y.-J. Lee and A.-L. Tan (Sense Publishers, Rotterdam, 2008), pp. 7–23.
Treagust, D. F. and Tsui, C.-Y., *Multiple Representations in Biological Education* (Springer, Dordrecht, 2013).
Triantafillou, C. *et al.*, Int. J. Sci. Math. Educ. **14**(4), 681–699 (2016).
Tsai, C. and Lydia Wen, M., Int. J. Sci. Educ. **27**(1), 3–14 (2005).
Tsaparlis, G., Chem. Educ. Res. Pract. **15**, 391–401 (2014).
Tural, G. *et al.*, J. Sci. Educ. Technol. **19**(5), 470–488 (2010).
United Nations, "General Assembly Resolution A/RES/70/1. Transforming Our World: the 2030 Agenda for Sustainable Development." (2015), see http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E (last accessed July 30, 2022)
Varvoglis, H., *History and Evolution of Concepts in Physics* (Springer, Cham, 2014).
Velentzas, A. *et al.*, Sci. Educ. **16**(3–5), 353–370 (2007).
Vojíř, K. and Rusek, M., Int. J. Sci. Educ. **41**(11), 1496–1516 (2019).
Vuković Stamatović, M., Corpus Linguist. Linguist. Theory **16**(3), 487–514 (2020).
Vybiral, B., *XXIX International Colloquium on the Management of Educational Processes, Pt 2*, edited by J. Neubauer and E. Hajkova (Univ Defence, Fac Econ & Management, 2011), pp. 551–560, see https://www.webofscience.com/wos/woscc/full-record/WOS:000394059500024.
Walford, G., Phys. Educ. **16**, 261–265 (1981).
Wang, T. *et al.*, Res. Sci. Technol. Educ. (published online 2021).
Weaver, J. F., J. Educ. Res. **39**(1), 42–55 (1945).
Wei, B. *et al.*, Sci. Educ. **31**(4), 943–960 (2022).
White, M. D. and Marsh, E. E., Library Trends **55**(1), 22–45 (2006).
Whiteley, P., Phys. Educ. **31**(3), 169–174 (1996).
Wilkinson, J., Res. Sci. Educ. **29**(3), 385–399 (1999).
Williams, H. T., Am. J. Phys. **67**, 670 (1999).
Winston, A. S. and Blais, D. J., Am. J. Psychol. **109**(4), 599–616 (1996).
Yun, E., J. Baltic Sci. Educ. **19**(3), 388–400 (2020).
Yuni, S. *et al.*, J. Phys.: Conf. Ser. **1811**, 012118 (2021).
Zajkov, O. *et al.*, Int. J. Sci. Math. Educ. **15**(5), 837–852 (2017).
Zhang, J. and Yang, K., Curriculum,Teaching Material and Method **42**(2), 138–144 (2022).
Zhuang, H. *et al.*, Int. J. Sci. Educ. **43**(11), 1779–1798 (2021).

CHAPTER 17

EVALUATION OF PHYSICS TEXTBOOKS

Marika Kapanadze, Gabriela Jonas-Ahrend, Alexander Mazzolini, and Fadeel Joubran

Kapanadze, M., Jonas-Ahrend, G., Mazzolini, A., and Joubran, F., "Evaluation of physics textbooks," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 17-1–17-30.

17.1 INTRODUCTION

The textbook (where "textbook" is defined to be the print version and/or its digital electronic equivalent) is recognized as an important source of knowledge whereby teaching materials are presented in a scientific but also intelligible way according to the curriculum being followed (Laketa and Drakulić, 2015). This chapter provides a broad review of research publications that focus on the evaluation of physics textbooks. It provides a systematic analysis of the internationally-available literature published in different journals, books, reports, and conference proceedings. To assist the reader, Fig. 17.1 shows an overview of the "key ideas" discussed in this review chapter.

While this chapter deals primarily with the evaluation of physics textbooks, there are some instances where this evaluation has been broadened to include a more general discussion on science textbooks. There are two reasons why this generalization is useful:

1. the discipline of physics is sometimes incorporated into science in several countries, especially at the secondary education level, and
2. there are results from many science textbook evaluation studies that are of direct relevance to physics textbooks.

Science textbooks play a critical role in learning and teaching at the primary, secondary and tertiary education levels. In 1996, Trowbridge & Bybee (as cited in Ogan-Bekiroglu, 2007, p. 599) noted that "Students generally feel more comfortable with a textbook than without one because textbooks emphasize important concepts, direct independent learning activities, and present learning goals for the study of a particular science topic." A study by Robinson *et al.* (2014) also noted "The textbook's role…

https://doi.org/10.1063/9780735425514_017

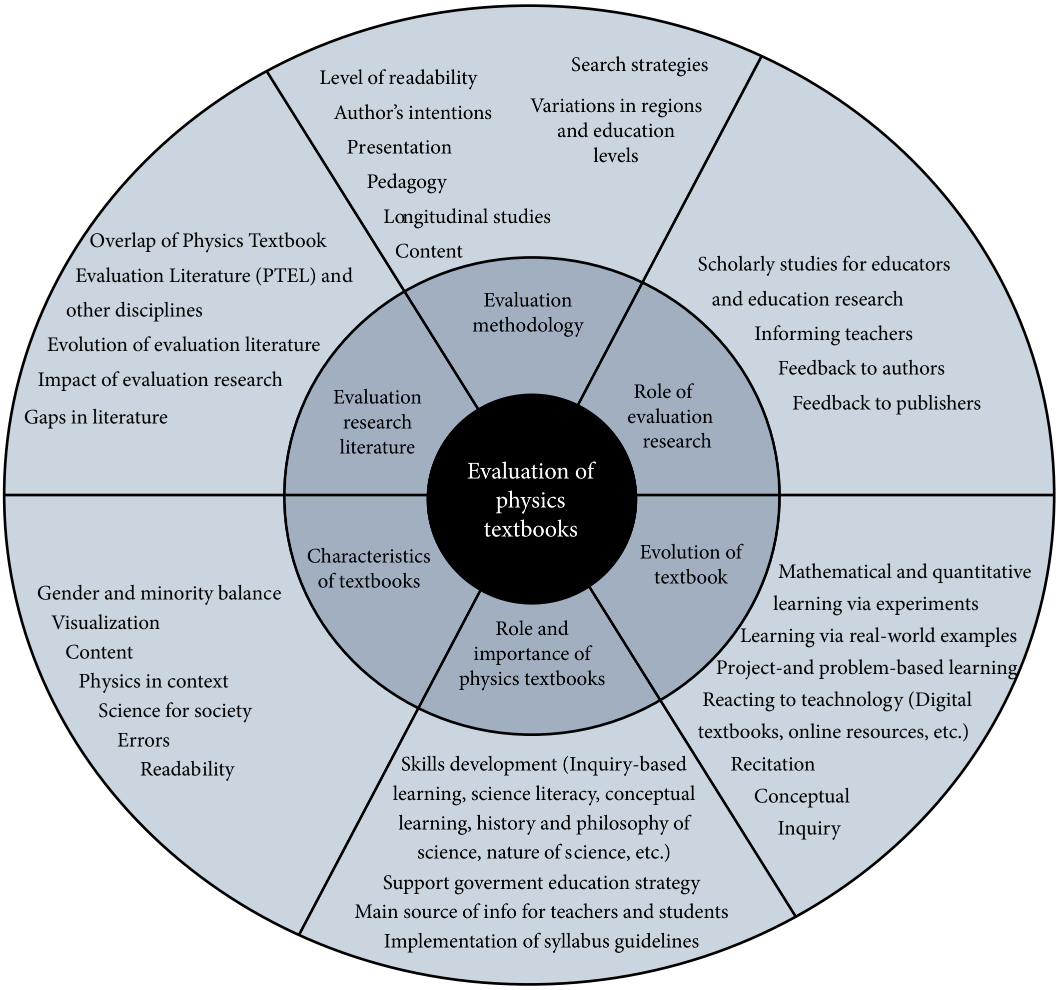


FIG. 17.1
Overview of the "key ideas" discussed in the review chapter.

extends beyond the dissemination of information. Textbooks play an important role in mediating the politics of what is taught, and even what methods are used to teach students."

With the increasing availability of digital learning materials over recent years, the use of digital textbooks has become common practice for many teachers and students in the U.S. (Ruggieri, 2020), and probably in many other western countries. Some online resources are not peer reviewed and may have physics errors; but others, such as the Khan Academy (https://www.khanacademy.org/), offer valuable alternative learning paths for students. These online education resources encourage multiple representations of physics ideas and content. They provide important opportunities for students to develop diverse learning frameworks.

In fact, the concept of the physics textbook has come full circle with the publication of online algebra-based and calculus-based textbooks that can be downloaded *for free* in PDF; for example, "College Physics" by Urone and Hinrichs (2020), and "University Physics Vol. 1" by Moebs *et al.* (2016), both published by OpenStax.

In the evaluation of textbooks, three domains (Reints, 2013) are usually identified: (1) content, (2) pedagogy, and (3) presentation. Content is important because it defines what needs to be studied and understood. But content must be transformed into something that students can understand and learn via pedagogy (i.e., the method and practice of teaching). In addition, deep learning must be supported and scaffolded by well-designed and well-presented resources.

One of the goals of this chapter is to discuss the effectiveness of Physics Textbook Evaluation Literature (PTEL) over the decades. This chapter discusses the evolution of PTEL (in terms of the types of research questions tackled, instruments employed, methodologies used, etc.), and the major turning points or milestones in its history. Based on published studies, this review investigates the following important questions relating to PTEL:

1. What are the most researched topics regarding content, pedagogy and presentation?
2. What topics are discussed in the most cited papers in this field?
3. Does PTEL favor any one specific education level?
4. Do any specific countries emphasize the evaluation of physics textbooks?

The review studies considered in this chapter differ in the time periods they cover. Most studies analyzed literature from several decades ago. Other studies were more recent. Dobricki *et al.* (2020) undertook a review concerning vocational learning and teaching using digital technologies and considered papers that had been published only in the previous five years. Vojíř and Rusek (2019) discussed textbook research trends in their review study, which analyzed 183 papers published over two decades.

Another review (on physics teaching and learning with multimedia applications) by Girwidz *et al.* (2019) analyzed 491 articles from 34 local-language journals from 10 countries, which must have been a complex undertaking. But in general, most reviews consider only (or mostly) literature that is written in English. For this review chapter, only English-language papers have been considered so that readers can easily refer to the original sources.

17.2 METHODS

17.2.1 Search strategy

The papers that are analyzed in this chapter were selected from different scientific databases and other online search engines. The chapter authors looked at the Web of Science, Scopus, ScienceDirect, Google

Scholar, and the reference lists of relevant scientific publications. Potentially interesting papers were searched using relevant keywords (e.g., physics textbooks, physics textbooks evaluation), and other important themes (e.g., scientific literacy, gender issues & equity, technology etc.). A total of 248 papers were selected for reading. Later, after discussions, the authors categorized relevant papers according to research topics and quality, and narrowed the final analysis to 126 papers.

17.2.2 Geographical and historical aspects

The review chapter takes a broad approach and includes the analysis of textbook-related papers from many countries across all six inhabited continents. It discusses the specific approaches and emphases of physics textbook evaluations used in various countries.

From a geographical perspective, this review found that the breakdown of the number of published English-language papers from each region is as follows – Europe: 34%, North America: 25%, Asia: 23%, Australia: 10%, Africa: 4%, and South America: 4%. The published physics textbook evaluation papers from Europe, North America and Asia are roughly comparable and account for approximately 80% of the total publications. Australia accounts for 10% of the publications, which is significant given its relatively low population. Africa contributes only 4%, most probably because many of the countries of Africa are classified as developing countries with limited education resources, and many different local languages. Interestingly, the review finds that only 4% of the published papers came from South America. This reflects a limitation of the review, which considers only PTEL published in English. The review authors suspect that the percentage of PTEL publications from South America is considerably higher but that these publications are written in Spanish or Portuguese (which makes the papers much more accessible to local academics and secondary-school teachers but excludes them from this review).

The distribution over time of the PTEL publication studied in this chapter is shown in Fig. 17.2. From this interesting historical perspective, the review found that the majority of PTEL papers have been published in the last decade. In this chapter review, 66 of the papers specifically focus on issues associated directly with physics textbooks, while the other 60 papers deal with fundamental issues that impact physics textbooks, such as the level of inquiry, online resources, and the nature of science.

17.2.3 Level of education

This chapter analyzes and discusses research papers that focus on the evaluation of physics textbooks from the primary to tertiary education level. Physics, as a separate subject, is taught from the upper secondary level in most of the countries in the world. Selected physics topics are also taught in primary-school science courses, and some examples of these are given in the chapter. This review also discusses the level to which science textbook evaluation studies favor primary, secondary, or tertiary education.

This review found the following breakdown for PTEL associated with the various education levels – primary: 14%, secondary: 59%, and tertiary: 27%. The review notes that PTEL associated with tertiary

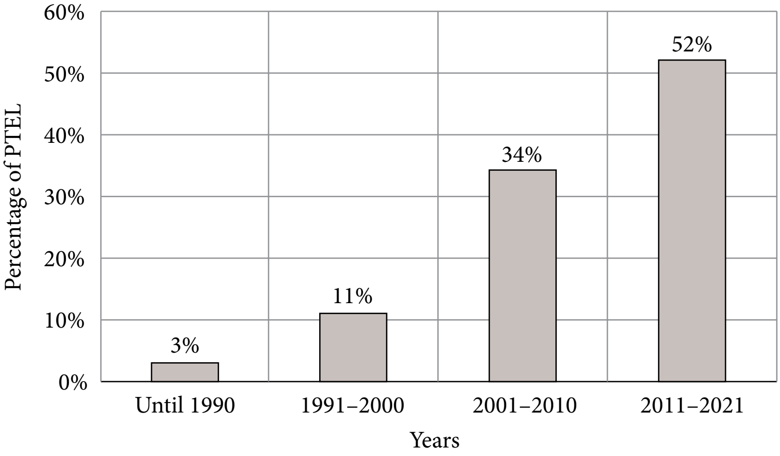


FIG. 17.2
The distribution over time of PTEL publication studied in this chapter.

studies is only about 27% of the total number of publications, compared to 73% associated with primary/secondary education studies. This imbalance perhaps indicates that primary- and secondary-school physics (and science) teachers rely on textbooks to a far greater extent than university academics.

17.3 FINDINGS

17.3.1 Development of physics textbook evaluation literature (PTEL)

This chapter discusses the development and evolution of PTEL and the major turning points and milestones of this development.

As mentioned earlier, the analyses of papers reviewed in this chapter are based on the three domains of Reints (2013), but in this case, the focus is on *how* these domains influence physics textbooks:

1. *Content:* This includes topics such as the textbook's relationship to technology, physics subject matter, physics taught in context, the nature of science (NOS), history & philosophy of science (HPS), gender balance, scientific literacy, readability, and the mistakes in textbooks.
2. *Pedagogy:* This includes topics such as active learning, the presentation of ideas through experiments, inquiry activities and their relationship to science skills development, and the use of analogies.
3. *Presentation*: This includes topics such as visualization, conceptualization and representation.

In addition to the three categories of Reints (2013), this review chapter investigates how physics textbooks are used by teachers and students, which includes the uses of textbooks in their different formats (printed and electronic).

17.3.1.1 Content

Papers in PTEL that investigate subject content consider either pedagogical issues (such as scientific literacy, the use of analogies, conceptual difficulties, thought experiments, collaborative learning, inquiry-based learning & teaching, and project-based learning) or educational issues (such as assessment & evaluation, the context in which physics is taught, history & philosophy of science, nature of science, online teaching, and curriculum).

Studies of physics textbooks cover a wide spectrum of subject-matter topics. Breaking down the distribution from 39 secondary-school physics textbook publications considered for this review chapter, it was found that 41% of the papers deal with Mechanics, 26% deal with electricity & magnetism, 18% deal with Optics & Modern Physics, and 15% deal with Laboratory work. Thus, it appears that Mechanics is the dominant subject area investigated by education research into secondary-school physics textbooks.

17.3.1.1.1 Nature of science (NOS) and history and philosophy of science (HPS)

NOS and HPS are important issues in science education research. It has been proposed that the integration of NOS and/or HPS into physics teaching could promote students' knowledge in optics (Galili, 2012) and other fundamental theories in physics (Galili, 2019). However, poor representations of NOS have been reported for some Chinese high-school physics textbooks (Li *et al.*, 2020), and similarly for some Greek textbooks (Kollas *et al.*, 2007). A study by Okoronka and Adeoye (2011) found that most Nigerian physics textbooks discussed very few HPS issues.

Abd-El-Khalick *et al.* (2017) analyzed 34 biology textbooks and 18 physics textbooks from the U.S.A. The research reported that, on average, less than 2.5% of the textbook pages were dedicated to addressing NOS constructs.

Niaz *et al.* (2010) analyzed 103 university textbooks that discussed HPS issues associated with the photoelectric effect. The study investigated six aspects that could be used to reconstruct the history of the photoelectric effect and rated them according to HPS aspects that were discussed in the textbooks. Four ratings were used for the comparison: "excellent," "satisfactory," "mention," and "no mention." The analysis revealed that only three of the six aspects obtained a score of "satisfactory," and none obtained a score of "excellent."

17.3.1.1.2 Technology

Today, technology is ingrained in the social lives of most students. They usually feel at ease with technology, which opens the door for using technological resources to enhance student learning and stimulate the development of cognitive skills. But there may still be a long way to go! A study by Souza and Garcia (2019) found that, for the "Modern Physics" sections of four Brazilian secondary-school physics textbooks, (1) most technological resources were passive requiring little engagement from the students, and (2) only one of the textbooks had technological resources that encouraged students to engage independently. Nevertheless, one teacher in the study commented that the use of technology in the classroom increased students' interest and that "Students begin to see Physics not just as mere mathematics." The Souza study concluded that in general, teachers need to adopt technological resources to stimulate students' interest in their physics lessons.

Van Nuland *et al.* (2020) noted that many STEM disciplines (like physics) are complex and abstract, and that they may benefit from technological tools that enhance learning through simulation and other online activities. During the Covid-19 pandemic, many courses had to operate in an online manner, which presented the opportunity of using technological tools to help facilitate experiential learning and to support learners' understanding of physical models and processes.

How, and to what extent the relationship between physics and technology is portrayed in physics textbooks is very important. In recent decades, there have been surprisingly few studies in PTEL that focus on the importance of technology in physics textbooks, even though technological aspects and applications are very common in physics textbooks. Gardiner (1999) analyzed five Canadian (university-level) physics textbooks and stated: "…if textbook authors continually present artifacts as nothing other than illustrations of physics principles in operation, then they will contribute very little to students' understanding of the nature of technology and may in fact promote the misconception that a knowledge of physics leads smoothly to the invention and production of useful technology. In real life, this is hardly ever true."

González *et al.* (2015) described their experiences with using smartphones to complement traditional learning in physics and found that smartphones had a "very positive influence on the students' engagement." They also used the smartphone sensors and interface to develop low-cost laboratory activities that were successful (if care was taken to ensure that the sensors' data were correctly accessed and interpreted).

A recent study by Yuni *et al.* (2020) investigated the amount of Science, Technology, Engineering, Art and Mathematics (often abbreviated as STEAM) found in five high-school physics textbooks from Indonesia. The study found that the textbooks generally did *not* emphasize the STEAM approach.

17.3.1.1.3 Gender balance

In the past decade, many formal barriers to women's participation in science and technology have been removed, but several informal barriers relating to gender, culture, and psychosocial effects remain. To better understand these phenomena, this review chapter focuses on what researchers, scientists and educators have identified as contributing to negative attitudes, low participation, and poor achievement in science for women. This chapter also investigates the role played by physics textbooks in the resulting female and minority attrition in the science pipeline.

In their extensive review of racial bias in physics textbooks from 1960 to 2016, Lawlor and Niiler (2020) found that women and minorities were still underrepresented in physics at nearly every level of education and academia. This troubling reality was reflected in every textbook they examined. The researchers concluded that, although there has been improvement, it has been very slow. Lawlor and Niiler suggested that it may be useful for male and white readers of their review paper to attempt a thought experiment in which all textbooks, which they study in a field, contain no images of anyone that looked like them. It would be hard to escape the conclusion that they could feel unwelcome or that they would feel that they did not belong in that field of study. Brotman and Moore (2008) described how inequities in the classroom, including gender-biased textbooks, resulted in boys and girls having different attitudes and levels of participation in science.

Sue (2010) wrote that "unintended, careless slights and inadvertent social cues take massive tolls on people from underrepresented groups, affecting both academic performance and psychological well-being." Pienta and Smith (2012) suggested that "textbook images also reflect a hidden curriculum that works to deter girls and women from academic and career interests in the fields of science, math or engineering." Blumberg (2008) suggested that gender bias in primary- and secondary-school books is "hidden in plain sight" and that "their stereotypes of males and females are camouflaged by the taken-for-granted system of gender stratification and roles which constrains girls' and boys' visions of who they are and what they can become."

A good example that tackles the lack of minorities in physics can be found in a paper by Nelson (2017) who suggested that minorities face similar struggles as women in science. The paper specifically pointed to covert and overt racism as major barriers in STEM fields and that these deep (direct and indirect) discriminations lead to low confidence, which when coupled to a lack of support and resources, resulted in a high dropout rate.

17.3.1.1.4 Scientific literacy

Scientific literacy (SL) is a term that has been used by educators since the late 1950s, and it can be defined as an individual's understanding of scientific concepts, phenomena and processes, and their ability to apply this knowledge to new situations. The importance of SL has been described in many educational research papers.

SL is associated with the process of thinking scientifically and identifying natural phenomena (Demir, 2016), and is also linked to critical thinking and collaboration skills (American Association for the Advancement of Science, 1989). The World Economic Forum (2015) identified 16 learning skills for the twenty-first century, and SL was included as one of those skills. A study by Karelina and Etkina (2007) suggested that the main purpose of current physics learning is to prepare the students for a better future in the 21st century.

Learning resources, and the education programs that support them are key factors that influence the SL of students, and their ability to develop SL skills. These skills are important as a preparation for students returning to the community after completing secondary school (Rusilowati *et al.*, 2016). Holbrook and Rannikmae (2009) suggested that SL has to be often taught from the viewpoint of "teaching through science" and not "science through teaching."

Some studies examined textbooks to determine the importance they place on SL. Rokhmah *et al.* (2017) investigated the topic of optical instruments in two secondary-school physics textbooks for grade 10 in Indonesia. Their analysis found that both textbooks taught most of the optical instrument topics through reading and writing activities only, and that explanations were transmitted directly, which did not give students the opportunity to engage in activities that helped them to construct their own understanding. Rokhmah's study concluded that "the learning process in school should support the development of students' science literacy skills, and that it can be done optimally by producing good physics textbooks to encourage students to do many kinds of activities such as experimental activities, observations, literature studies, as well as role playing." Rokhmah and his colleagues suggested that SL skills could help students to solve problems scientifically and in an accountable manner.

A quantitative analysis (Wilkinson, 1999) examined SL themes used in senior secondary-school physics textbooks (from the State of Victoria, Australia, between 1967 and 1997). These themes were part of the Victorian Certificate of Education (VCE) physics curriculum, and the syllabus was detailed in the VCE Study Design. Wilkinson's analysis showed that:

1. the majority of the textbooks emphasized the importance of science as a "body of knowledge,"
2. science as a "way of investigating" appeared to be of less importance,
3. very little emphasis was given to science as a "way of thinking," and
4. the textbooks used for the *revised* VCE physics course (post 1990) paid more attention to the theme of science, technology and society compared to the course before 1990.

A study by Kollas *et al.* (2007) analyzed Greek physics textbooks. The purpose of the study was to examine the content of the officially-approved physics textbooks of Greek lower secondary education regarding their level of SL compared to the stated general objectives of the Greek curriculum. The analysis followed the method developed by Chiappetta *et al.* (1991a, 1991b). The Study by Kollas indicated that "knowledge of science" was the most discussed aspect of SL, at the expense of the other aspects of SL that were defined in the Greek curriculum. In general, 68.5% of the SL elements were identified for "knowledge of science," 22.6% for the category of "investigative nature of science," 7% for the category of "interaction among science, technology, and society," and only 1.9% for the category of "knowledge about science." While knowledge of science (i.e., presenting current theories, models, facts, concepts and laws) was well represented in the Greek textbooks, the second category, "investigative nature of science," was underrepresented and did not support students in understanding the scientific process itself. Again, there appeared to be little support in Greek textbooks for students to develop an understanding of "knowledge about science" and "interaction among science, technology, and society." The study (Kollas *et al.*, 2007) concluded that the textbook authors needed to pay more attention to these aspects.

A study by Tairab (2005) examined SL in UAE secondary-school science textbooks. Tairab chose nine physics textbooks for the analysis (three for each of the 1st, 2nd and 3rd-year of secondary school). The study analyzed four aspects of SL, which were based on a previous categorization used by the American Association for the Advancement of Science (1993) (Boujaoude, 2002; and Stern and Roseman, 2004). Tairab's study showed that the main emphasis in all the textbooks analyzed was on "scientific knowledge," which was the first aspect of SL (56%–66% for each grade). The other three SL aspects had lower percentages: "investigative activities," which was the second aspect (23%–30% for each grade), "a way of knowing," which was the third aspect (7%–10% for each grade), and "a way of impacting societies," which was the fourth aspect (1%–7%, for each grade). Tairab concluded that there was a need for more communication between educators and textbook authors to consider the results of the study, and that decision-makers should also be aware of these results for future textbook selection.

The analysis of physics textbooks regarding SL skills has also been undertaken by international programs such as PISA (Programme for International Student Assessment). Recent PISA analysis confirmed the results of the study by Kollas *et al.* (2007) and showed that Greek students had greater difficulties in the application of scientific knowledge in real life when compared to the other students from participating countries.

The physics textbooks recommended by the Ministry of Education for Turkish secondary schools (9th and 10th grades) were studied by Türk *et al.* (2018), who used the PISA science literacy competence levels assessment and evaluation activities. Türk's study concluded that there was a need to increase the quality of the assessment and evaluation activities in the reviewed Turkish textbooks. The implementation of Türk *et al.* (2018) recommendations may assist students to improve their ability to create new knowledge, analyze new information, and interpret situations connected to everyday life, as outlined in the SL component of the PISA science framework.

17.3.1.1.5 Mistakes in textbooks

Physics textbooks provide a foundation for teaching and learning, and ideally, they should be without flaws from both the pedagogical and factual points of view. However, a lot of Physics Education Research (PER) papers reveal that textbook mistakes (or errors) can encourage physics misconceptions for the learner (and the inexperienced teacher) at all levels of study – primary and secondary school, and even in tertiary education. In fact, textbook errors may also reflect the misconceptions of textbook authors, and these misconceptions can be passed on to their readers. Hence, conducting research to reveal the mistakes or even ambiguities in textbooks plays an important role in improving students' understanding of physics.

Slisko (2009) compared the "culture of teaching" to the "culture of research" in the discipline of physics. Slisko found that the biggest difference was that errors in research publications were corrected *before* publication through the mechanism of peer review used in research journals. However, the lack of meaningful peer review in many textbooks means that textbook errors are published, and that they can remain for many years.

Research reveals that mistakes in physics textbooks occur in many topics such as mechanics, electricity & magnetism, optics, fluid mechanics, and modern physics. A few examples are given in the following paragraphs.

Anselmo *et al.* (2020) analyzed how Pascal's principle was discussed in 9 undergraduate textbooks from Brazil. The study reported that there were inconsistencies regarding the authorship of the principle, the concept of pressure, the definition of a fluid, and the application of the principle. The study compared the textbooks with the original work of Pascal and found "inconsistencies" both from the conceptual and historical points of view.

Gezerlis and Williams (2021) discussed several errors in "numerical methods and computational physics," which were common in several higher education textbooks. In their paper, the authors discussed the mistakes and provided corrections and illustrations to clarify the errors.

Zajkov *et al.* (2017) analyzed an 8th grade physics textbook that was written in the Macedonian language and then translated to other languages. Mistakes were found in both didactic and subject-matter areas. Zajkov's analysis focused only on three topics: electric current, magnetic fields, and electromagnetic induction. One example from Zajkov's study (p. 844) propagated a known misconception in the relationship between current and voltage. The textbook had the following statement: "In order to obtain higher voltage (and higher current) it is necessary to move the magnet faster, use stronger magnet and use coil with more loops." This statement could form the common misconception that if there is a voltage, there should be a current. This is not always correct.

Another study by Yildiz (2015) focused on the expression "All quantities can be converted into length, time and mass units in the end. No matter how complex any physical quantity is, it is stated as an algebraic combination of these three basic quantities." This statement was found in several university physics textbooks. The study correctly concluded that the statement could be considered true only for mechanics. Yildiz recognized that there are *seven* fundamental quantities that can be used to express all other quantities found in engineering and science. The seven fundamental quantities are length, mass, time, electric current, temperature, number of moles, and luminous intensity (LMTAΘNC).

Several studies have investigated the concept of "weight" in physics textbooks (Galili and Lehavi, 2003; and Taibu *et al.*, 2015). These studies concluded that different, inconsistent, or ambiguous uses of the concept in different physical situations could lead to conceptual difficulties for some students in secondary schools.

Atoms, which are a highly abstract concept, are presented in physics textbooks via various models. According to Hejnová and Králík (2019), eight different models have been found in Czech textbooks for lower secondary schools. The models of atoms differ in color, the structure of the electron shells, electronic orbits and relative sizes of nuclei. While textbook authors are free to choose the atomic model that best suits their needs, some models have the potential to encourage misconceptions among students. For example, the coloring or shading of atomic components may encourage the idea that all spaces in the atom are filled with something or, if the model includes electron shells, students could be misled towards the misconception that removing electrons from the atom can break a hard shell.

In geometric optics textbooks, images can be either virtual or real. The eye of the observer plays an important role in the formation and observation of these images. Gürel and Eryilmaz (2013) analyzed 10 physics textbooks from around the world to determine the role of the observer's eye in image formation and found that it was ignored in virtual images in plane mirrors, and real images in lenses. Hence, there is a possibility that readers of these textbooks could develop misconceptions in their interpretation of geometric optics.

Definitions of concepts such as heat, energy, weight and electric current have significant implications with respect to physics education. The lack of accuracy and coherence about these definitions in physics textbooks may lead to misconceptions for teachers and students. Accurate definitions in textbooks can prevent such misconceptions or learning misdirection (Galili and Lehavi, 2006). A study by Develi and Namdar (2019) looked at the definitions of friction force in 26 Turkish physics and science textbooks and found several inconsistencies in their explanations of friction force. Develi & Namdar proposed a more comprehensive definition that could minimize student misconceptions.

17.3.1.1.6 Readability

Reading is one of the most frequently used forms of science instruction reported by teachers; therefore, students must have the ability to interpret the print-based information that they read. Reading is viewed as the active construction of a text's meaning, involving an interaction between the writer and reader. Yet, even well-written textbooks are of little use if students do not have the knowledge and skills to effectively use them. Education research shows that teaching students to recognize and use the organization and structure of text increases their reading comprehension considerably (Penney, 2000). In fact, Koch (2001) described the development, application, and evaluation of a metacognitive technique to improve students' reading comprehension of physics texts. Koch's experimental study with 64 students in a pre-university, one-year introductory physics course showed that the metacognitive treatment was an effective tool for promoting students' reading-comprehension ability. A recent review (Singer and Alexander, 2017) concluded that the medium used (print or screen) may also play an important role in text readability and comprehension.

17.3.1.1.7 Physics in context

"Context-based learning" involves teaching a discipline (in this case physics) within the framework of a "real-world" environment with which students are already familiar. Teaching in context can help motivate students to be more engaged in their own learning. As mentioned earlier (Wilkinson, 1999), the Victorian Certificate of Education (VCE) secondary-school physics curriculum was redesigned (around the early 1990s) so that physics "areas of study" were embedded within broad themes of students' everyday experiences. The context determined which physics content areas were covered and in what order (Whitelegg and Parry, 1999). The contextual approach was adopted by several secondary-school textbooks, for example, “Physics: Revealing our world, Book 2” (Mazzolini *et al.*, 1992). Whitelegg and Parry found that after the introduction of the new physics curriculum, there was a significant increase in the number of students taking the VCE physics course.

17.3.1.2 Pedagogy

17.3.1.2.1 Historical aspects of textbooks and the inclusion of laboratory work

The content and teaching styles of most physics textbooks have ebbed and flowed throughout history. Calinger *et al.* (2019) reported that up to the mid-seventeenth century, the forerunners of modern physics textbooks were based on the Aristotelian teachings from Ancient Greece. These textbooks did not employ mathematics, did not make references to experiments and did not investigate scientific problems. The chief purpose of these textbooks was to confirm truths. The practice in these early teachings of science was to read the text, which was the authority, and discuss and argue its main principles using deduction.

The Aristotelian point of view (Davidson, 1892) was challenged and largely surpassed later in the seventeenth century, mainly by the French philosopher and scientist René Descartes, who used a deductive method that grounded physics in mathematics, and the English statesman Francis Bacon, who applied an inductive empirical method. Eighteenth-century textbooks in Europe and North America continued with this “Cartesian” philosophy and, as they became more “scientific,” started to describe experimental apparatus and mechanisms.

A review by Meltzer and Otero (2015) reported on the historical development of physics education in the U.S.A., and its influence on physics textbooks. Their study can be summarized as follows:

1. Prior to the 1850s, instruction in science (including the discipline of “natural philosophy”) was strongly connected to textbooks, and recitation (i.e., students reciting from their textbooks) was the normal practice.
2. In the 1850s, the qualitative nature of textbooks started to change to a more quantitative approach (via practice problems at the end of chapters).
3. In the late 1800s, some textbooks started to include a few laboratory experiments as part of an inductive teaching method (and textbooks outlined experiments that students could undertake before the class).
4. In the early 1900s, practical experiments still played a minor role in physics textbooks. These experiments were prescriptive and designed to illustrate known laws and principles.
5. Between the first and second world wars, textbooks began to focus on ways to teach physics via real-world applications.
6. During the “Cold War” period, the importance of strategic technological advancement meant that physics education emphasized fundamental unifying principles and a calculus-based approach. The Physical Science Study Committee (PSSC), and subsequent similar national science committees developed new curricula and textbooks to support these new directions.

7. Over the last 50 years or so, the widespread dissemination and acceptance of PER has meant that decisions concerning changes in teaching and learning approaches have been guided by scientific data and reasoning rather than personal beliefs or "gut instinct." Over that period, there has been a significant diversity in the approaches of different physics textbooks.

Meltzer & Otero (2015) suggested that most of the educational changes that occurred over the past 150 years or so had been in response to the personal views of influential educators and physicists or to national education committees that attempted to interpret the state of education in the U.S.A. Most textbooks tried to respond to these education changes, and they cycled between traditional Cartesian teaching and inquiry-based learning approaches.

Following the education reforms of the 1960s and 70s, the role of practical experiments in physics textbooks changed, and there was an increased emphasis on the laboratory as a basis for thinking and learning about science, and on developing skills through inquiry (e.g., observing, recording, drawing logical conclusions, etc.) rather than on simple verification of laws and principles (Saunders, 1992). Consequently, physics textbooks often integrated laboratory work into their educational offerings. The role of the laboratory was strengthened during this period because curriculum developers started to produce complete packages of materials, equipment, and guides, which often included a separate workbook for laboratory activities.

Laboratory activities are an important aspect of physics learning as they require students to investigate and study the concepts and laws of physics to understand the natural world (Hofstein and Lunetta, 2004). These activities are central to the learning process and students can conduct them individually or collaboratively in laboratory groups. Hence, it can be argued that there is some merit in linking laboratory activities to the textbook used by the students (Gumilar and Ismail, 2021).

A study by Bryant (2006) reported that unguided, inquiry-based laboratory investigations resulted in knowledge gains that were greater than those from traditional laboratory methods. Inquiry-based laboratory activities remain an essential aspect of physics learning, and their integration into mainstream textbooks and workbooks should provide opportunities for improving the learning process (Penney *et al.*, 2003).

But physics textbooks do not necessarily always follow education best practices even when those practices are embedded in the curriculum. A recent study of laboratory activities in physics textbooks was conducted by researchers from Finland and the United States (Park and Lavonen, 2013). They examined the questioning style and level of inquiry activities in two popular, comparable, high-school physics textbooks (one from each country) to determine how agreed reforms in the curriculum aligned with textbook laboratory activities. While both curricula emphasized the importance of open-ended, inquiry-based experimental activities, neither textbook provided laboratory activities at a high level of open inquiry. All experimental activities in both textbooks were determined to be at a very low level (i.e., where the problem and procedures were fully specified and only the solution was left open). Fortunately, this study also found that students using these textbooks were provided with other opportunities to practice inquiry problem-solving skills during the experiments, including using tools to gather, analyze, and interpret data, and proposing solutions, explanations, predictions and communicating the results. These skills can be acquired only when students engage in "active inquiries."

Another study of some Ethiopian physics secondary-school textbooks (grade 9–12) was undertaken by Assefa (2020), who investigated whether the textbooks made suggestions on what practical experiments could be undertaken by students in their physics courses, and how well the suggested experiments were integrated with the theory discussed in the textbook. Assefa found that attempts were made by the textbook writers to include many experiments in every chapter, but that the experiments were viewed as being subordinate to the theory (rather than being an essential part of it). Most experiments were described by Assefa as being "theory illustration and verification without giving opportunities for students to construct the scientific meanings of concepts."

Tamir (1976) reported on the use of laboratory activities in physics teaching over the past few centuries. While educators' views concerning the importance of laboratory activities varied considerably over that period, Tamir suggested that if students do not understand the concepts embedded in the scientific investigation processes, they will fail to develop problem-solving skills that are a key component of laboratory activities. Tamir's report proposed that science courses should give students many laboratory investigation opportunities so that they can practice and acquire scientific skills around concepts, problem-solving, and attitudes. As laboratory activities are often embedded in modern textbooks and their associated laboratory manuals, they should have a positive influence on the development of these skills. In this manner, students will be able to nurture the cognitive and analytical skills that are normally associated with rigorous scientific investigation.

Over the past 50 years or so, evidence from PER has encouraged the development of inquiry-based physics courses. During the same period, computer-based laboratory technologies have been developed and utilized in guided inquiry. In some physics courses, student engagement, both in small group teaching and in large "active learning" lecture classes, has been actively pursued. Herron noted in 1971 (as cited in Yang and Liu, 2016, p. 2690) that during this period, several significant national curriculum reforms had occurred in the U.S. and other countries, and that several inquiry-based textbooks had been developed and adopted; however, independent evaluation of these textbooks had often lagged.

Emphasis in textbooks can vary from country to country. An analysis of electric circuit lessons in both a Finnish and a Thai science textbook revealed that the Thai textbook emphasized procedural knowledge, while conceptual knowledge was emphasized mostly in the Finnish textbook (Sothayapetch *et al.*, 2013).

It is also interesting to note that several popular textbooks have not fundamentally changed during the latter half of the 20th century, but rather that they have added additional materials to address some PER findings. In her honors thesis, Stewart (2006) compared two editions of the popular university-level physics textbook Halliday and Resnik. She compared the 3rd edition published prior to the development of the Force Concept Inventory (FCI), which is a test to quantitatively measure students' conceptual understanding of mechanics and dynamics, and the 7th edition published after the FCI. Stewart found little substantial change in emphasis (at least in the chapters on mechanics), even though the two editions were published almost two decades apart. The two editions were very similar in scope and direction, though the 7th edition was more mathematical in its approach. The newer edition was more attractive and made better use of diagrams than its predecessor, but was probably slightly less readable and more verbally complex. The 7th edition did alert readers to some physics misconceptions (highlighted by FCI research) via "key ideas" and "cautions," but Stewart concluded that the two editions were very similar in scope.

In a reflection by Hewitt (1995), he speculated that trying to teach physics concepts via problem-solving activities was expeditious but not very effective. In the 1990s, Hewitt's textbook "Conceptual Physics" (Hewitt, 1989) became popular. It focused on explaining the concepts of physics and took a less mathematical approach than traditional physics textbooks. While aimed primarily at courses for non-physics majors, many educators have integrated parts of Hewitt's textbook into their own physics courses.

During the first two decades of the 21st century, PER has continued to develop, and it acts as a catalyst for a constructivist, inquiry-based approach to physics instruction in many teaching institutions. A lack of experience with scientific inquiry may hinder students' ability to gain a deep understanding of scientific ideas. Consequently, scientific inquiry is a highly recommended learning process in which students propose ideas based on evidence derived from their practices, develop understandings about scientific concepts and make sense of how to engage in science (Yang and Liu, 2016).

Many science textbooks have been criticized for presenting science as a complete body of information that has been derived without setbacks or errors (Chiappetta *et al.*, 1991a), and that this type of presentation may hamper physics students' abilities to construct meaning and develop an appropriate understanding of the knowledge, nature, and processes of science (Penney, 2000). A study by Glynn and Muth (1994) noted that "students who are learning constructively will challenge the science text they are reading or writing, struggle with it and try to make sense of it by integrating it with what they already know."

The national science curriculum reforms around the turn of this century (in the U.S.A. and other countries) have further strengthened the constructionist framework of science education. The report of the National Science Education Standards (1996) in the U.S. suggested that textbooks should include more activities that promote students' active involvement in higher-order thinking. This recommendation has led to the introduction of several inquiry-based physics textbooks, which have a very different emphasis compared to traditional textbooks. Examples of physics textbooks that have attempted to include an inquiry-based instruction are "Active Physics" (Eisenkraft, 1998), "PhysicAL: An activity approach to physics" (Martin, 1994), "Physics by Inquiry" (McDermott *et al.*, 1995), "Six Ideas That Shaped Physics" (Moore, 1997), and "RealTime Physics: Active Learning Laboratories" (Sokoloff *et al.*, 2011).

17.3.1.2.2 Analogies

Students' deeply-rooted misconceptions can affect their learning from physics textbooks. Students can misinterpret the physics behind their real-world experiences, and this may hinder their comprehension of physics textbooks. Certainly, learning from text is a complex undertaking involving the knowledge and interests of the learner, characteristics of the text, and the features of the context (Alexander and Kulikowich, 1994).

Analogies are sometimes used in physics textbooks to clarify physics concepts (and misconceptions). Unfortunately, these analogies may not always facilitate comprehension. A study by Didiş and Hidir (2019) analyzed six Turkish science textbooks used in primary and secondary schools (grades 3 to 8). The study evaluated the use of analogies as a teaching strategy using the "teaching-with-analogy" approach of Glynn (1994, 2007). The results of the analysis reveal that most of the analogies used in primary/secondary-school science teaching relate to physics concepts. However, 54.1% of them were rated as "poor," 44.3% were rated as "moderate," and 1.6% were rated as "good." Didiş and Hidir presented two examples of textbook analogies that were rated "moderate" from the topics of gravitational force (7th grade) and electrical conductivity (6th grade). Using Glynn's approach, the researchers presented revised versions of these two analogies that would probably prevent students' misconceptions to a greater extent than the original versions.

In their historical analysis of electric current in textbooks from 1891 to 1991, Stocklmayer and Treagust (1994) noted that in the textbooks they had studied, there had been "remarkably little change in their presentation of direct-current circuitry, most texts by implication portraying, in various ways, a fluid model which predates Faraday."

17.3.1.3 Presentation

17.3.1.3.1 Visualization and conceptualization

Several studies have compared and evaluated physics textbooks with a focus on the textural treatment of a particular topic or content.

Larkin (1983) found many differences in the problem-solving performance of experts and novices using different visual representations. Given their importance, relatively little is known about the variety and the efficiency of visual representations in physics textbooks.

Bungum (2008) analyzed the changing character of visual images in a sample of nine Norwegian physics textbooks from 1943 to 2008. Images were analyzed according to the categories of Dimopoulos *et al.* (2003), which included content specialization, social-pedagogic relationships (framing), and level of image abstraction (formality). A set of five modes of imaging physics was constructed and presented in a historical perspective. Bungum's group concluded that "While textbooks from the first half of the

20th century to a high degree presented both scientific objects and experiments in realistic ways as a foundation for knowledge in physics, newer textbooks tend to present the same subject matter with a high content specialization by means of conceptual entities and generalized models." The study also investigated how images communicate physics to the learner and how this communication, as well as the learner's role, has changed over the years.

Several textbook studies from Ethiopia, Indonesia and Brazil have provided a broader analysis of visualization. They considered the interplay between text and visualizations, a research analysis on textbooks, and thought experiments in textbooks. The results from three of these studies are summarized as follows – Zewdie (2014) analyzed two physics textbooks (grade 7 and 8) from eight Ethiopian upper primary schools from the perspective of six categories: learning objectives, text narratives, activities, figures & diagrams, review questions & problems, and unit summaries. The study included document analysis (coding of the textbooks) and open-ended questionnaires from 12 physics teachers and 80 students. The study reported that all six categories were found in each chapter, though there were differences in quantity and quality.

Bancong and Song (2018) analyzed 30 physics textbooks from Indonesia and focused on thought experiments. The researchers stated that "The study concludes that Indonesian physics textbooks published from 2009 to 2017 generally lack thought experiments. Many authors of these Indonesian physics textbooks ignored or inadequately presented thought experiments. …. So, in general, thought experiments presented in the Indonesian physics textbooks cannot be used as an introduction in transferring scientific knowledge to science students."

In 1985, the Federal Government of Brazil released its National Textbook Programme (PNLD), which regulated the production, evaluation, selection and distribution of textbooks for Brazilian primary and secondary public schools. A study by Santos *et al.* (2019) investigated a number of Brazilian textbooks by analyzing academic education publications from 2009–2017. They selected six academic journals which were available online and found 65 papers that claimed a research link to the physics textbooks. Out of these, they selected 16 papers for deeper analysis. The study outlined eight categories for research on physics textbooks: constitution of the textbook, environmental education, experimentation, science history, para-didactic books, problem solving, imaging representations, and didactic transposal. Their analysis gave an overview of the areas of education and physics teaching that were being investigated for Brazilian textbooks. They concluded that educational research on Brazilian physics textbooks was very limited, and that it did not reach 1% of the total publications in the highest-rated journals in Brazil, even though the Federal Government made big investments in the evaluation of textbooks. The most researched category in the 16 papers was didactic transposition, while the history of science was not investigated. One paper discussed how "imaging-verbal representation contribute (sic) to the understanding of Coulomb's torsion balance, concluding that the representations do not yield understanding of the concept of this experimental apparatus."

For many decades, studies from Shavelson (1972) and Merzyn (1987), and others have analyzed the terminology used in physics textbooks. For example, Merzyn stated that German physics textbooks "overwhelm students with too many different terms." Härtig (2014) also analyzed the terminology within German lower-secondary physics textbooks and concluded that "it is critical to analyze how the content is presented to the students," but unfortunately his paper only considered formulas and no other representations.

A study by Yun (2020) analyzed the linguistic differences in terminology for the concept of "force" between a popular high-school physics textbook (Holt Physics by Serway & Faughn) and a widely-used university introductory physics textbook (Fundamentals of Physics by Halliday and Resnick). The linguistic analysis only considered written text and did not include illustrations. Yun found that about 70% of the text in the university textbook had new words that had not been used in the high-school textbook.

As previously mentioned in this chapter, a study by Hejnová and Králík (2019) analyzed images of atoms in physics textbooks from the Czech Republic. The study concluded that very different models exist in textbooks, both from the didactic point of view and from graphic design.

17.3.1.3.2 Virtual and augmented reality technologies

In addition to printed and digital textbooks, virtual and augmented reality technologies (VR and AR) are sometimes used for teaching and learning in primary/secondary schools and universities. Kravtsov and Pulinets (2020) described a model of a learning system using AR technologies for visualizing illustrations in a secondary-school (grade-8) physics textbook in the Ukraine. Their interview study of 16 secondary-school teachers "showed the possibility, interest and effectiveness of using electronic learning results," and their survey of STEM students showed that students had a "willingness to work with AR technologies."

Two Indonesian studies (Bakri *et al.*, 2019; and Mahardika *et al.*, 2020) described the development of physics textbooks with AR/VR technologies. After a formative evaluation of the media, materials, and learning feasibility of the technologies, Bakri's study concluded that textbooks equipped with AR technology are appropriate and feasible. The study by Mahardika focused on students' higher order thinking skills, which can be nurtured by inquiry-based textbooks accompanied by multiple representations.

Online learning, used as an extension of the traditional laboratory workbook, can also provide a different perspective to traditional experimental activities. For instance, augmented virtual laboratories allow students to easily redo the same activity so that they can learn from their mistakes. They can also conduct technically-demanding experiments in different online environments or investigate simulated experiments that could not be conducted in a real laboratory. A review study by Brinson (2015) found that students' learning achievements were similar or greater for virtual laboratory activities compared with traditional hands-on laboratories.

17.3.1.4 Use of printed textbooks and online/multimedia resources

17.3.1.4.1 Printed textbooks

There are many studies that have evaluated the use of physics textbooks by teachers and students. Before the internet, textbooks were often the *only* resource readily available to students, and many teachers considered textbooks as the sole teaching resource (Maffia *et al.*, 2003).

The appropriateness of physics textbooks for teaching and learning has been evaluated by Ogan-Bekiroglu (2007), who noted that "Textbooks do not only influence what and how students learn but also what and how teachers teach." Ogan-Bekiroglu developed an instrument for identifying the characteristics of high-school physics textbooks and used it to analyze Turkish textbooks. She concluded that "textbooks approved by the Ministry [of Education in Türkiye, formerly known as Turkey] do not meet the criteria supporting the effective physics teaching and learning (at the secondary school level)." Indeed, many Turkish secondary-school teachers experience difficulties when trying to use these approved textbooks for promoting student inquiry. Ogan-Bekiroglu noted that teachers use physics textbooks "as means of imparting factual knowledge." However, she stressed the importance of inquiry-based learning, and how it should have a positive influence on the development of future textbooks. The evaluation analysis of Ogan-Bekiroglu was not only useful for the Turkish Ministry of Education but also for physics teachers and textbook evaluators from the rest of the world, and for the selection of suitable textbooks in other science disciplines.

In general, the evaluation of textbooks helps teachers to select the most appropriate book for their situation and gives feedback to authors and publishers on how textbooks can be improved. Some studies (e.g., Leite *et al.*, 2013) discuss the role of physics textbook evaluation in pre-service teacher preparation so that future teachers can be better prepared to make important choices regarding the textbooks they will use. Leite's study investigated physics teacher-training courses in southern Brazil and showed that student teachers' knowledge about the differences between textbooks was inadequate. About 70% of teachers participating in the study did *not* think that physics degree courses prepared student teachers to adequately evaluate, select and use the textbooks provided by Government programs. The study concluded that in the future, student teachers should be better trained on how to assess and select textbooks, and how to understand the different ways of using textbooks. In 2001, a study by Thompson *et al.* (as cited in Ogan-Bekiroglu, 2007, p. 602) suggested that when deciding on the best textbook to use, teachers should evaluate the difficulty of the textbook, the reading level of their students, the students' motivation level, their workload, and the cost of the textbook.

The use of physics textbooks as a tool to support collaborative learning was investigated by Boxtel *et al.* (2000). Fifty-six students from the Netherlands (aged between 15 and 16) participated in the study. They were randomly assigned to the same gender pairs within each class, as boys were dominant in mixed gender groups in science classes. The use of textbooks during concept mapping tasks and their consequent influence on students' interaction was investigated. The study by Boxtel found that textbooks were consulted for an average of 33% of the total time spent on the learning tasks. Students used the textbooks most frequently when they started a new part of a task. While one student was preparing for the task (writing or reading), the other student in the pair would consult the textbook (so as not to disturb the other student). Some students had difficulty finding information as they had less experience in using textbooks and no clear idea of what they were looking for. The study found that the students often only skimmed the text. Frequently, they would focus on text with bold printing or on colored backgrounds. The table of contents was rarely used. When students elaborated from the textbook, they did so in their own words and without discussion or collaboration with their partner. Boxtel's study concluded that the use of textbooks resulted in a constrained, elaborative interaction, but that more collaboration was observed in situations without textbooks.

Kalman *et al.* (2008) suggested that studying a physics textbook for understanding is often difficult for students because there can be a significant gap between their prior knowledge and the conceptual knowledge demands of the subject. The researchers noted that often students didn't read the textbook in conjunction with classroom activities. Instead, they preferred to use the textbook as an aid in solving problems by finding what they perceived to be relevant solved problems or other useful information.

Cummings *et al.* (2002) investigated the use of student textbooks in introductory physics courses (for future scientists and engineers at two U.S. education institutions – a polytechnic institute, and a university). The study focused on two aspects – (1) how the placement of solved examples influenced students' use of the textbook, and (2) the utilization of course assignments for encouraging students to read the textbook. The Cummings study found that 35%–45% of the students did *not* use the textbooks at all during the semester, and that many students who were reading the textbooks found the dispersed solved example most helpful. The study also reported that the level of student reading was remarkably low at the polytechnic institute. The main conclusion of these researchers was that "students have not figured out for themselves that reading is a potentially useful intellectual endeavour," and they suggested that teachers should encourage students to read more so that they can become self-directed learners.

The use of textbooks for pre-class assignments in Canadian university physics and biology courses was studied by Heiner *et al.* (2014). The physics course was an introductory, calculus-based course for science majors. Students reported that pre-reading had a positive effect on their general learning and class preparation. Heiner's study found that 80% of students read the textbooks regularly. The results were very similar for the biology course and textbooks.

17.3.1.4.2 Digital textbooks

Will digital textbooks (and more generally online multimedia resources) ever replace the printed textbook? In 1913, Thomas Edison (as cited in Reints, 2013, p. 29) predicted that "books will soon be obsolete in the schools… …It is possible to teach every branch of human knowledge with the motion picture. Our school system will be completely changed in ten years." Reints suggests that the reason this has *not* happened is that all these new media, from motion pictures to mobile phones, are not primarily developed as teaching devices. All these digital resources have great potential, but their implementation needs to be based on the results of education research.

Digital textbooks can vary text size, highlight important passages, and allow searches of relevant words and phrases; however, the choice of paper or screen may play an important role in a textbook's readability and a student's comprehension. A detailed review by Singer and Alexander (2017) found that comprehension was similar when students read short texts from paper or screen, but that comprehension decreased when long texts were read from a screen (rather than from paper) due to the increased cognitive load associated with scrolling.

In recent decades, the widespread development and availability of the internet and the World Wide Web (WWW) has meant that many students now seek information from numerous different sources rather than solely from the prescribed textbook.

But online learning does not work for *all* students. A study by Knight (2015) found that students who preferred using electronic versions of textbooks found it easier to navigate digital content, but that other students struggled with navigating through, and working in, an online space. The Knight study revealed that some teachers found it quite difficult to find both time and support to master online materials and to successfully integrate them into their teaching.

While technology has had an important impact on student behaviors and preferences for learning, one thing that has not changed is the essential need for credible content. Technology is useless without valid content. Useful educational technologies have reliable content, support students to interact with the materials and engage students in their own learning. Some online resources are not peer reviewed and may have physics errors, but many of them, such as the Khan Academy, offer valuable alternative and complementary learning approaches. These online education resources encourage multiple representations of physics ideas and content, and provide many opportunities for students with diverse learning frameworks.

Ruggieri (2020) investigated students' perceptions and their use of online education resources as a supplement to the materials and activities already provided in courses. Ruggieri used online surveys and interviews for his study of 1st and 2nd year physics courses at a U.S. university. In addition to the prescribed textbook, free online media resources such as YouTube, Khan Academy, and a fee-based online tutoring/textbook rental website (Chegg) were used during the course. The way in which students in Ruggieri's study acquired their textbooks is shown below:

- 48.8% purchased the cheaper online version of the prescribed textbook
- 13.5% purchased a new print copy of the prescribed textbook
- 11.3% did not obtain either the paper or online version of the prescribed textbook
- 6.9% borrowed the textbook from someone else
- 6.3% used teacher-provided excerpts of the textbook

- 3.0% purchased a used paper textbook
- 3.0% used the free “OpenStax” online non-prescribed textbook
- 1.9% rented the textbook
- 0.3% obtained a copy of the prescribed textbook from the library
- 5.1% either used a different method of textbook usage or did not specify details

(*Note that the data add to a total of 100.1%, which is due to rounding issues.*)

Cost was a significant factor in how students acquired their textbooks (56% of students reported that cost was very important, 31% reported that it was somewhat important, and only 13% reported that it was not important).

Ruggieri found that for the 1st-year students, 30%–47% used the prescribed textbook occasionally or often, but that 53%–70% used the prescribed textbook never or rarely. Ruggieri also found that the 2nd-year students used their textbooks somewhat more than the 1st-year students. The only unique textbook purpose reported by students was to summarize key points. In contrast, Ruggieri showed that most students used online resources regularly as part of their weekly study cycle. In fact, 88%–92% of 1st-year students, and 97% of 2nd-year students reported using online resources often or occasionally, rather than never or rarely. The analysis showed that students mostly use online resources rather than any form of textbook. The two most popular online resources used by 1st- and 2nd-year students were YouTube (76%–88%) and Khan Academy (63%–79%). In fact, only 5% of all students used *no* online resources, while 79% used between one and four different online resources, and 16% used five or more different online resources.

Ruggieri found that students mostly used course and online resources for completing homework tasks, and that only a few of them sought out in-person resources (due to time constraints). Students used online resources (such as YouTube and Khan Academy) to help them develop conceptual understanding and problem-solving skills as well as guided practice with problems. Online textbook repository services were used to access line-by-line solutions to assessment problems. Students reported difficulties with using textbooks and found that they did not need them to complete the required course tasks. Students also mentioned that the prescribed textbook was not well integrated into the course structure.

Online resources are used by many students because they facilitate quick access to a range of learning styles. Multimedia modules are often presented in short, focused activities that allow students to quickly engage with confusing concepts. When information is transmitted traditionally to the learner through the textbook, the learner is viewed as a passive recipient. Engaging students with interactive multimedia modules can promote experiential learning. For example, in a study by Stelzer *et al.* (2009), the learning of basic physics content for undergraduate students who used a typical contemporary introductory textbook was worse than that of students who used multimedia modules that were designed using principles developed from research into multimedia learning.

Another study (Pol *et al.*, 2005) of physics textbooks compared (amongst other things) the relationship between the acquisition of problem-solving skills and the choice of learning materials. Using pre- and post-tests, the study compared the achievements of two lower-secondary student groups. Pol's study concluded that the group that was taught with both the textbook and a computer-supported tool achieved better results (in developing problem-solving skills) than the control group that used only the textbook.

A decade-long, U.S. study (Seaton *et al.*, 2014) found that the structure of a course has a strong influence on how much, and when, students accessed electronic textbooks during their coursework. Physics courses that deviated strongly from traditional assessment structures, most notably by more frequent exams and other regular assessments, show consistently high usage of the on-line materials with far less "cramming" before exams.

While care must be taken not to overload students with too much choice, multimedia modules offer many different learning modalities (including video, simulations, text, pictures and sound), which can be easily adapted to the learning style, motivation and cognitive capabilities of individual students. In addition, online activities enable the easy implementation of diverse learning strategies including assessment of "prior knowledge" and provision of detailed and timely feedback (Reints, 2013).

It could be argued that the widespread availability of online resources is challenging the prescribed textbook as the predominant study tool of physics students. In fact, the concept of purchasing a physics textbook has come full circle with the publication of free, online textbooks; for example, "College Physics" by Urone and Hinrichs (2020) published by OpenStax.

After the declaration of the Covid-19 pandemic in March 2020, most educational institutions throughout the world have had to move away from face-to-face teaching and adopt wide-scale online learning and remote content-delivery (at least for the duration of the pandemic). This has meant that teachers and students have had to rapidly adapt to synchronous and asynchronous learning via online activities and to integrate these activities into their institution's learning management systems. While physics and other STEM disciplines are well suited to the use of digital resources to enhance students' learning, the various e-activities need to be grounded in good physics education research, which has not always been the case because of the unexpected rapid shift to online learning (due to the Covid-19 pandemic).

17.4 CONCLUSIONS

The evaluation of textbooks is critically important as over 90% of all science teachers traditionally have tended to rely entirely on textbooks for their teaching (Park, 2005), and hence, textbooks play a very important role in students' learning. In a 2007 Pakistan study, Mohammad and Kumar (as cited in Sothayapetch *et al.*, 2013, p. 59) found that the science teachers in rural communities had difficulties in recognizing mistakes in textbooks and in using textbooks to optimize the learning process. Textbook authors have a responsibility to ensure that their textbooks are written clearly, have information that is accurate and relevant, and that all errors have been identified and eliminated. Physics textbooks are written by scientists, educators, or teachers who may be guided by their prior epistemological beliefs (Kahveci, 2010) rather than by science education research. Another U.S. study (Hubisz, 2003) showed that the most popular physics and science textbooks used in secondary schools at the turn of the 21st century were content-wise inaccurate, filled with errors, and hence unacceptable from learning and teaching perspective.

Gönen and Kocakaya (2006) surveyed high-school physics teachers in Türkiye (formerly known as Turkey) and found that most physics textbooks were perceived as insufficient in terms of scientific content. According to the teachers in their study, this insufficiency was due to the scarcity of solved physics problems, lack of alignment between the physics curricula and the National University Examinations, and the perception that textbook content was outdated when compared to modern developments in science and technology.

There has been considerable research into the evaluation of science textbooks in recent decades (Khine, 2013) – some studies have mostly addressed language analysis, images, analogies, and textual features (Dimopoulos and Karamanidou, 2013; and Muspratt and Freebody, 2013), while others have focused on the investigation of diagrams, gender, and themes of scientific literacy (Zohar and Sela, 2003; Elgar, 2004; Liu and Treagust, 2013; and Park and Lavonen, 2013).

Very few investigations have specifically focused on how scientific methodology is represented in textbooks (see, for example, Blachowicz, 2009) and these studies have been somewhat limited (i.e., just one part of the investigation, or just a few chapters of the textbook, or undertaken from a philosophical (rather than scientific) perspective (Binns, 2013).

The review by Vojíř and Rusek (2019) of 183 papers published between 2000 and 2018 showed that researchers in Europe and the U.S.A. focused on textbook research to a greater extent than researchers elsewhere in the world. The most frequently researched books are science textbooks for secondary schools. Textbook research consists mostly of analyzing learning concepts and how they are integrated, non-textual elements in textbooks, visual representations, learning content, or learning text analysis. The number of research papers per year focused on the analysis of science textbooks has increased over the past few decades from an average of around 3–4 in the early 2000s to an average of around 20 in the late 2010s. But overall, the number of research papers focused on science textbooks is low compared to the number of scientific research papers in total (average of around 0.0005% from 2000 to 2018) (Vojíř and Rusek, 2019).

17.4.1 Final remarks

This chapter explores the development of physics textbook evaluation literature (PTEL) and includes the contribution of PTEL to the current knowledge of physics textbooks from the viewpoint of content, pedagogy and visualization, as well as PTEL's limitations and level of influence in textbook development. The number of papers published in PTEL has increased significantly over the past decade. This promising trend may lead to improvements in physics textbooks, and physics education more generally.

Some gaps in PTEL have also been identified. For example, in sustainable development goals (SDG), topics such as "climate change," "clean energy," "water resource management" perhaps should be included in physics textbooks. Classroom discussions about these topics could bring students to a deeper understanding of the problems affecting our world and how physics and technology can contribute to alleviating climate change and sustainability issues.

Another gap in PTEL is the relationship between the disciplines of Science, Technology, Engineering and Mathematics (STEM), and the discipline of Art. Perhaps a more holistic approach in STEM may prove helpful, and the inclusion of aspects of artistic endeavor may help physics students to learn in a more creative way.

There needs to be more research into the evaluation criteria used to assess the quality of textbooks (Ogan-Bekiroglu, 2007). Secondary-school teachers and university academics need to have the evaluation skills to choose the best textbooks for their courses. As discussed earlier (Leite *et al.*, 2013), secondary-teacher training programs do not equip student teachers with sufficient skills for textbook evaluation, and this hinders the ability of new physics teachers to select appropriate textbooks in their everyday school life.

In general, there are few studies on the extent to which PTEL results influence the development of textbooks. Many academics probably do not focus their attention on PTEL. Most secondary-school teachers probably do not read scientific publications and journals, and very rarely attend scientific conferences. Conference organizers often offer reduced registration fees for teachers to engage and motivate them to participate in scientific discussions. It is very important to share PTEL results via teacher journals and conferences. Linking PTEL with teachers and academics will help close the gap between secondary and tertiary levels of education, researchers and teacher practitioners, and will support the general improvement of physics teaching.

This review concludes that while PTEL serves a useful purpose in stimulating further research into the evaluation of physics textbooks, in general, the results of PTEL do not have sufficient influence on the positive development of physics textbooks.

Feedback to textbook authors is also important. One of this chapter's contributors (AM) was a co-author of two successful upper secondary-school physics textbooks in Australia. As mentioned earlier in the chapter, the first textbook, entitled "Physics: Revealing our world, Book 2" (Mazzolini *et al.*, 1992), was written from the context of "physics explained via everyday experiences." AM contributed four chapters on electronics to this textbook. Electronics was explored through designing and constructing several hands-on projects exploring a variety of electronic aids for use with families, the elderly, and the disabled. The "in context" emphasis in the textbook aligned with the State Government guidelines on the syllabus and the requirement to present physics concepts within the context of everyday life.

The second textbook, entitled "Heinemann Physics 12" (Chapman *et al.*, 2004), was a more traditional physics textbook, again written to follow the State Government guidelines on the syllabus at the time (which emphasized the core concepts of traditional physics and introduced several modern "cutting-edge" topics as electives). This new approach used core units to cover traditional physics topics and a series of elective units to cover more modern concepts that were designed to engage and excite students. In this textbook, AM wrote the core units on electronics and introductory photonics, and the elective "detailed study" on "photonics and fibre optics."

Although the two physics textbooks were very different in both scope and contextualization, the author groups of both textbooks worked hard to embrace the detailed guidelines developed by the State Education Department for the two very different syllabi, and to reflect the syllabus guidelines in the textbooks. AM acknowledges that, sadly, he did not read any research-based textbook evaluation studies before writing his contributions to the two textbooks. He simply tried to faithfully interpret the syllabi and guidelines, and used his teaching experience, knowledge of physics, and understanding of PER to help him write in a coherent, factual and hopefully engaging manner. Similarly, it appears that other textbook authors may only occasionally investigate PTEL.

It may also prove useful if physics textbook users (mainly teachers) could give more feedback to authors and perhaps provide reports about their experiences. Students are indirect receivers of the benefits of PTEL. This review chapter did *not* find papers where physics textbook evaluation involved *both* students and teachers. It would be interesting to see if a future study of this kind could show whether students and teachers evaluate physics textbooks similarly, or whether there are variances in how they perceive different aspects such as content, design, and learning support.

ACKNOWLEDGMENTS

The authors of this chapter thank Professor Pamela Fraser-Abder for her valuable contribution.

REFERENCES

Abd-El-Khalick, F. *et al.*, J. Res. Sci. Teach. **54**(1), 82–120 (2017).

Alexander, P. A. and Kulikowich, J. M., J. Res. Sci. Teach. **31**(9), 895–911 (1994).

American Association for the Advancement of Science (AAAS), *Science for all Americans* (Oxford University Press, 1989), see http://www.project2061.org/publications/sfaa/online/sfaatoc.htm.

American Association for the Advancement of Science (AAAS), *Benchmarks for Science Literacy* (Oxford University Press, 1993), see http://www.project2061.org/publications/bsl/online/index.php.

Anselmo, D. H. A. L. *et al.*, Eur. J. Phys. **41**(6), 1–21 (2020).

Assefa, S., IOSR J. Human. Soc. Sci. **25**(1), 57–64 (2020).

Bakri, F. *et al.*, J. Penelit. Pengemb. Pendid. Fis. **5**(2), 113–122 (2019).

Bancong, H. and Song, J., J. Pendid. IPA Indonesia **7**(1), 25–33 (2018).

Binns, I. C., *Critical Analysis of Science Textbooks: Evaluating Instructional Effectiveness*, edited by M. S. Khine (Springer, Dordrecht, 2013), pp. 239–258.

Blachowicz, J., British J. Philos. Sci. **60**, 303–344 (2009).

Blumberg, R. A., Prospects **38**(3), 345–361 (2008).

Boujaoude, S., Int. J. Sci. Educ. **24**(2), 139–156 (2002).

Boxtel, C. V. *et al.*, J. Exp. Educ. **69**(1), 57–76 (2000).

Brinson, J. R., Comput. Educ. **87**, 218–237 (2015).

Brotman, J. S. and Moore, F. M., J. Res. Sci. Teach. **45**(9), 971–1002 (2008).

Bryant, J. A., J. College Sci. Teach. **35**, 56–61 (2006). (Note: Original article was published with incorrect author name, R. Bryant).

Bungum, B., NorDiNa **4**(2), 132–141 (2008).

Calinger, R. S. *et al.*, *Leonhard Euler's Letters to a German Princess* (Morgan and Claypool Publishers, 2019), Chap. 1, pp. 1–18.

Chapman, R. *et al.*, *Heinemann Physics 12*, 2nd ed. (Heinemann Harcourt Education, 2004).

Chiappetta, E. L. *et al.*, J. Res. Sci. Teach. **28**(8), 713–725 (1991a).

Chiappetta, E. L. *et al.*, J. Res. Sci. Teach. **28**(10), 939–951 (1991b).

Cummings, K. *et al.*, *[Paper Presentation] Physics Education Research Conference 2002*, Boise, Idaho, 7–8 August 2002 (AAPT, 2002), see https://www.compadre.org/Repository/document/ServeFile.cfm?ID=4374&DocID=1097.

Davidson, T., *The Great Educators*, edited by N. H. Butler (Charles Scribner's Sons, 1892), see https://archive.org/details/aristotleancient00davirich.

Demir, E., Int. Educ. Stud. **9**(4), 99–107 (2016).

Develi, F. and Namdar, B., J. Educ. Sci. Environ. Health **5**(1), 91–101 (2019).

Didiş (Körhasan), N. and Hıdır, M., Phys. Rev. Phys. Educ. Res. **15**(1), 1–8 (2019).

Dimopoulos, K. and Karamanidou, C., *Critical Analysis of Science Textbooks: Evaluating Instructional Effectiveness*, edited by M. S. Khine (Springer, Dordrecht, 2013), pp. 61–78.

Dimopoulos, K. *et al.*, Res. Sci. Educ. **33**, 189–216 (2003).

Dobricki, M. *et al.*, Int. J. Res. Vocat. Educ. Train. **7**(3), 344–360 (2020).

Eisenkraft, A., *Active Physics*. It's About Time. Internet Archive (1998), see https://archive.org/details/activephysics00eise.

Elgar, A. G., Int. J. Sci. Educ. **26**(7), 875–894 (2004).

Galili, I., Sci. Educ. **21**(9), 1283–1316 (2012).

Galili, I., Sci. Educ. **28**(3), 503–537 (2019).

Galili, I. and Lehavi, Y., Am. J. Phys. **71**(11), 1127–1135 (2003).

Galili, I. and Lehavi, Y., Int. J. Sci. Educ. **28**(5), 521–541 (2006).

Gardiner, P., Int. J. Sci. Educ. **21**(3), 329–347 (1999).

Gezerlis, A. and Williams, M., Am. J. Phys. **89**(1), 51–60 (2021).

Girwidz, R. *et al.*, Int. J. Sci. Educ. **41**(9), 1181–1206 (2019).

Glynn, S., Sci. Child. **44**(8), 52–55 (2007).

Glynn, S. M., *Teaching Science with Analogy: A Strategy for Teachers and Textbook Authors. Reading Research Report No. 15* (National Reading Research Center, 1994), see https://eric.ed.gov/?id=ED373306.

Glynn, S. M. and Muth, D., J. Res. Sci. Teach. **31**(9), 1057–1073 (1994), see https://studylib.net/doc/8376201/reading-and-writing-to-learn-science.

Gönen, S. and Kocakaya, S., J. Turk. Sci. Educ. **3**(1), 40–42 (2006).

González, M. Á. *et al.*, J. Cases Inform. Technol. **17**(1), 31–50 (2015).

Gumilar, S. and Ismail, A., Res. Sci. Technol. Educ. (published online 2021).

Gürel, D. K. and Eryilmaz, A., Hacettepe Üniv. Eğitim Fakültesi Dergisi **28**(2), 234–245 (2013), see https://fdocuments.net/document/a-content-analysis-of-physics-textbooks-as-a-probable-a-content-analysis.html?page=1.

Härtig, H., *Science Education Review Letters* (Humboldt-Universität zu Berlin, 2014), pp. 1–7.

Heiner, C. E. *et al.*, Am. J. Phys. **82**(10), 989–996 (2014).

Hejnová, E. and Králík, J., AIP Conf. Proc. **2152**(1), 1–6 (2019).

Hewitt, P. G., *Conceptual Physics*, 6th ed. (Harper Collins, 1989).

Hewitt, P. G., Phys. Today **48**(9), 85 (1995).

Hofstein, A. and Lunetta, V. N., Sci. Educ. **88**(1), 28–54 (2004).

Holbrook, J. and Rannikmae, M., Int. J. Environ. Sci. Educ. **4**(3), 275–288 (2009), see https://files.eric.ed.gov/fulltext/EJ884397.pdf.

Hubisz, J., Phys. Today **56**(5), 50–54 (2003).

Kahveci, A., Int. J. Sci. Educ. **32**(11), 1495–1519 (2010).

Kalman, C. *et al.*, J. College Sci. Teach. **37**(4), 74–81 (2008).

Karelina, A. and Etkina, E., Phys. Rev. Spec. Top. **3**(2), 020106 (2007).

Khine, M. S., *Critical Analysis of Science Textbooks Evaluating Instructional Effectiveness* (Springer eBooks, 2013), see https://link-springer-com.ezproxy.lib.swin.edu.au/book/10.1007%2F978-94-007-4168-3.

Knight, B. A., Cogent Educ. **2**(1), 1–10 (2015).

Koch, A., Sci. Educ. **85**(6), 758–768 (2001).

Kollas, S. *et al.*, *[Paper Presentation] IOSTE International Meeting, "Critical Analysis of School Science Textbooks,"* Yasmin Hammamet, Tunisia, 7–10 February 2007 (IOSTE, 2007).

Kravtsov, H. and Pulinets, A., *ICTERI Workshops* (ICTERI, 2020), pp. 918–933, see http://ceur-ws.org/Vol-2732/20200918.pdf.

Laketa, S. and Drakulić, D., Interdiscipl. Descript. Complex Syst.: INDECS **13**(1), 117–127 (2015).

Larkin, J. H., *Mental Models*, 1st ed., edited by D. Gentner and A. L. Stevens (Lawrence Erlbaum Associates, London, 1983), pp. 75–98.

Lawlor, T. M. and Niiler, T., Phys. Teach. **58**(5), 320–323 (2020).

Leite, A. E. *et al.*, *Textbooks and Educational Media in a Digital Age. Proceedings of the Thirteenth International Conference on Research on Textbooks and Educational Media*, edited by Z. Sikorova *et al.* (IARTEM, 2013), pp. 118–125, see https://iartemblog.files.wordpress.com.

Li, X. *et al.*, Res. Sci. Educ. **50**(3), 833–844 (2020).

Liu, Y. and Treagust, D. F., *Critical Analysis of Science Textbooks: Evaluating Instructional Effectiveness*, edited by M. S. Khine (Springer, Dordrecht, 2013), pp. 287–300.

Maffia, A. M. C. *et al.*, *[Paper Presentation]. European Science Education Research Association (ESERA) Conference,* Noordwijkerhou, Netherlands, 19–23 August 2003 (ESERA, 2003).

Mahardika, I. K. *et al.*, J. Phys.: Conf. Ser. **1465**(2020), 1–9 (2020).

Martin, B., *PhysicAL: An Activity Approach to Physics*, 2nd ed. (J.M. LeBel Enterprises, 1994).

Mazzolini, M. *et al.*, *Physics: Revealing Our World, Book 2* (The Jacaranda Press, 1992).

McDermott, L. C. *et al.*, *Physics by Inquiry—An Introduction to Physics and the Physical Sciences* (John Wiley & Sons, 1995), Vol. 1.

Meltzer, D. E. and Otero, V. K., Am. J. Phys. **83**(5), 447–458 (2015).

Merzyn, G., Int. J. Sci. Educ. **9**(4), 483–489 (1987).

Moebs, W. *et al.*, *University Physics* (Rice University, OpenStax, 2016), Vol. 1, see https://openstax.org/books/university-physics-volume-1/pages/1-introduction.

Moore, T. A., *Six Ideas That Shaped Physics* (McGraw-Hill College, 1997).

Muspratt, S. and Freebody, P., *Critical Analysis of Science Textbooks: Evaluating Instructional Effectiveness*, edited by M. S. Khine (Springer, Dordrecht, 2013), pp. 33–60.

National Science Education Standards, *National Committee on Science Education Standards and Assessment* (National Research Council, 1996), see http://www.nap.edu/catalog/4962.html.

Nelson, A., Phys. Today **70**(5), 10–11 (2017).

Niaz, M. *et al.*, Sci. Educ. **94**(5), 903–931 (2010).

Ogan-Bekiroglu, F., J. Sci. Teach. Educ. **18**(4), 599–628 (2007).

Okoronka, A. U. and Adeoye, F. A., Euras. J. Phys. Chem. Educ. **3**(2), 75–83 (2011), see https://www.ijpce.org/index.php/IJPCE/article/view/115.

Park, D. Y., J. Geosci. Educ. **53**(5), 540–547 (2005).

Park, D. Y. and Lavonen, J., *Critical Analysis of Science Textbooks: Evaluating Instructional Effectiveness*, edited by M. S. Khine (Springer, Dordrecht, 2013), pp. 219–238.

Penney, K., M.Ed. thesis, Faculty of Education, Memorial University of Newfoundland (Memorial University Research Repository, 2000), see https://research.library.mun.ca/9034/.

Penney, K. *et al.*, Canad. J. Sci. Math. Technol. Educ. **3**(4), 415–436 (2003).

Pienta, R. S. and Smith, A. M., *The New Politics of the Textbook. Constructing Knowledge (Curriculum Studies in Action)* (Sense Publishers, 2012), Vol. 1.

Pol, H. *et al.*, Int. J. Sci. Educ. **27**(4), 451–469 (2005).

Reints, A., *Textbooks and Educational Media in a Digital Age. Proceedings of the Thirteenth International Conference on Research on Textbooks and Educational Media, 18–20 September 2013*, edited by Z. Sikorova *et al.* (IARTEM, 2013), pp. 15–31, see https://iartemblog.files.wordpress.com.

Robinson, T. J. *et al.*, Educ. Res. **43**(7), 341–351 (2014).

Rokhmah, A. *et al.*, J. Pendid. Fis. Indonesia **13**(1), 19–24 (2017).

Ruggieri, C., Phys. Rev. Phys. Educ. Res. **16**(2), 1–25 (2020).

Rusilowati, A. *et al.*, J. Pendid. Fis. Indonesia **12**(2), 98–105 (2016).

Santos, T. A. *et al.*, *Researching Textbooks and Educational Media From Multiple Perspectives: Analysing the Texts, Studying Their Use, Determining Their Impact. IARTEM 2019, 15th International Conference on Research on Textbooks and Educational Media*, edited by S. T. Gissel (UCL University College, Laeremiddel.de—The Danish National Centre of Excellence for Learning Resources, Odense, 2019), pp. 51–61, see https://laeremiddel.dk/wp-content/uploads/2020/08/IARTEM-2019-Proceedings-Final_re.pdf.

Saunders, W. L., School Sci. Math. **92**(3), 136–141 (1992).

Seaton, D. T. *et al.*, Am. J. Phys. **82**(12), 1186–1197 (2014).

Shavelson, R. J., J. Educ. Psychol. **63**(3), 225–234 (1972).

Singer, L. M. and Alexander, P. A., Rev. Educ. Res. **87**(6), 1007–1041 (2017).

Slisko, J., Phys. Commun. Cooperation **2**, 31–46 (2009), see http://physics.le.ac.uk/girep2009/ConferenceProceedings/GIREP2009_ConferenceProceedings_Volume2.pdf.

Sokoloff, D. R. *et al.*, *Real Time Physics: Active Learning Laboratories, Module 1: Mechanics* (John Wiley & Sons, 2011).

Sothayapetch, P. *et al.*, Eur. J. Math., Sci. Technol. Educ. **9**(1), 59–72 (2013).

Souza, J. L. L. and Garcia, T. M. F. B., *Researching Textbooks and Educational Media From Multiple Perspectives: Analysing the Texts, Studying Their Use, Determining Their Impact. IARTEM 2019, 15th International Conference on Research on Textbooks and Educational Media*, edited by S. T. Gissel (UCL University College, Laeremiddel.de—The Danish national Centre of Excellence for Learning Resources, Odense, 2019), pp. 72–82, see https://laeremiddel.dk/wp-content/uploads/2020/08/IARTEM-2019-Proceedings-Final_re.pdf.

Stelzer, T. *et al.*, Am. J. Phys. **77**(2), 184–190 (2009).

Stern, L. and Roseman, J. E., J. Res. Sci. Teach. **41**(6), 538–568 (2004).

Stewart, B. J., Honors in physics dissertation (University of Arkansas, 2006), see https://www.researchgate.net/publication/237507228.

Stocklmayer, S. M. and Treagust, D. F., Sci. Educ. **3**(2), 131–154 (1994).

Sue, D. W., *Microaggressions in Everyday Life: Race, Gender, and Sexual Orientation* (John Wiley & Sons, New York, 2010), see https://books.google.co.il/books?hl=en&lr=&id=jyzcuvgTaIMC&oi=fnd&pg=PR9&ots=pAKme5YrPT&sig=fKrmB4oRwlOxqkcB7hEWul5GC8k&redir_esc=y#v=onepage&q&f=false.

Taibu, R. *et al.*, Phys. Rev. Spec. Top. **11**(1), 1–20 (2015).
Tairab, H., Int. J. Book **3**(2), 31–37 (2005).
Tamir, P., "The role of the laboratory in science teaching," Technical Report Series, Vol. 10, pp. 1–33 (Science Education Center, Iowa University, 1976), see https://eric.ed.gov/?id=ED135606.
Türk, O. *et al.*, Eur. J. Educ. Stud. **5**(1), 42–55 (2018).
Urone, P. P. and Hinrichs, R., *College Physics* (Rice University, OpenStax, 2020), see https://openstax.org/details/books/college-physics.
Van Nuland, S. E. *et al.*, FASEB BioAdv. **2**(11), 631–637 (2020).
Vojíř, V. and Rusek, M., Int. J. Sci. Educ. **41**(11), 1496–1516 (2019).
Whitelegg, E. and Parry, M., Phys. Educ. **34**(2), 68–72 (1999).
Wilkinson, J., Res. Sci. Educ. **29**(3), 385–399 (1999).
World Economic Forum, *New Vision for Education: Unlocking the Potential of Technology* (British Columbia Teachers' Federation, Vancouver, BC, 2015).
Yang, W. and Liu, E., Int. J. Sci. Educ. **38**(18), 2688–2711 (2016).
Yildiz, A., Eur. J. Phys. Educ. **6**(3), 1–7 (2015), see https://www.proquest.com/scholarly-journals/discussion-on-expression-written-about/docview/1807489954/se-2.
Yun, E., Eur. J. Phys. **41**(6), 065704 (2020).
Yuni, S. *et al.*, J. Phys.: Conf. Ser. **1811**(1), 012118 (2020).
Zajkov, O. *et al.*, Int. J. Sci. Math. Educ. **15**(5), 837–852 (2017).
Zewdie, Z. M., Am. J. Educ. Res. **2**(1), 44–49 (2014).
Zohar, A. and Sela, D., Int. J. Sci. Educ. **25**(2), 245–268 (2003).

SECTION IV MATHEMATICS IN TEACHING AND LEARNING PHYSICS

Section Editor

Gesche Pospiech

This section is dedicated to a central point of physics as a science, and thus to the nature of physics and the role of mathematics. The use of mathematics in physics has been discussed from many perspectives throughout history. From a philosophical point of view, the fascination that a science such as mathematics, which deals with ideal objects and is characterized by rigorous logical deduction and proof, is suited to describe natural processes so thoroughly and accurately, prevails. During the development of physics, the importance of mathematics for physics has steadily increased and brought about numerous advances. In light of the long and persistent investigations and discussions of the relationship between mathematics and physics, it is somewhat surprising that this topic became an independent research topic in PER relatively late. Except for a few papers and some approaches in the problem-solving literature, research in the context of PER began to intensify only around the year 2000. In the relatively short time since then, considerable progress has been made in both theoretical foundations and empirical research, although research desiderata remain. Above all, it has become clear that the role of mathematics in physics and the view of pupils and students in this interplay and their handling of it are now seen in a much more differentiated way than before. This has made it possible to provide initial hints for the design of instruction in physics education, both in school and at the university.

Therefore, in Chapter 18, “Role of mathematics in physics from multiple perspectives,” first the philosophical and historical perspectives are briefly presented. The main part is then devoted to the perspective of physics education and the importance and role of mathematics in the educational context at school or university. In order to describe mathematization, i.e., the gradual translation of physical facts into mathematical language with the help of mathematical elements or structures, different points of view are taken. On the one hand, the distinction between technical and structural dimensions proves to be fruitful. Furthermore, the roles of mathematics in physics from the point of view of teaching and learning can be identified: the reduction of cognitive load by equations or formulas, the objectivity, the exactness and the facilitation of communication by mathematics. This mathematization goes hand in hand with developing and using models. The necessary blending of physical and mathematical perspectives, which is a necessary precondition for mathematization, problem solving, and gaining new insights, is difficult for most students. A deeper explanation of difficulties and the following strategies need theoretical contributions from other fields of knowledge that address cognitive possibilities and strategies of learners in mathematization or provide background theories, which are described.

Equations and formulas play a particularly important role in physics—also from the perspective of learners and teachers. Therefore, a separate Chapter 19, “The meanings of physics equations in the context of physics education,” is devoted to this large and important area. Here, once again, the relationship between physics and mathematics is briefly presented with special reference to equations in both mathematics and physics. Emphasis is placed on working out the physical meaning of equations by verbalizing them. Appropriate teaching-learning activities are described for this purpose and the role of conceptual knowledge is discussed. The related processes are analyzed using the framework of ontological categories and epistemic classification of equations. The chapter is rounded off by describing empirical results on learning difficulties, both at the student and teacher levels. These studies relate both to students’ multiple problems in understanding equations and solving problems, but also to the use of successful strategies. It turns out that students’ epistemic beliefs play an important role in their performance during mathematization. A variety of difficulties were also reported at the university level. In particular, it was shown that successful completion of tasks is not necessarily associated with conceptual understanding. Here, the choice of a favorable instructional method seems to remedy the situation, at least in part. The contribution further elaborates that attention to mathematization is also needed in teacher education so that teachers become aware of the complexity of the interplay of mathematics and physics and are able to identify student difficulties.

Besides equations, graphs, especially line graphs, play an important role because of the functional relationships that are central in physics. These are discussed in Chapter 20, “Graphics.” Graphs are one of the most important forms of representation for doing physics, communicating about physics and also for learning physics. Research has shown that dealing with graphs is not as easy as one might think, but that numerous cognitive demands must be overcome in reading, in interpreting, as well as in constructing them. As a result, significant difficulties, some of them specific, arise in both mathematics and physics. Despite decades of research, questions remain open, for example about students’ problems or strategies in constructing graphs. This contribution presents and discusses different diagnostic instruments that can assess graphing skills at different educational levels. Important new insights are also expected from new research methods such as eye-tracking. Research on this topic has also shown that observed problems are not always due to a lack of mathematical knowledge. Rather, it seems that the transfer between mathematics and physics education needs to be actively promoted. It is expected that this also applies to other areas of mathematization, such as the handling of formulas.

Especially in the area of graphical representations, microcomputers have also made early inroads and have been used, especially in kinematics. More recently, advances in computing technology have opened up significant opportunities to use digital tools in physics education. This is discussed in Chapter 21, “Visualization and mathematization: How digital tools provide access to formal physics ideas.” The authors bring together two fields, visualization and mathematization, and study how digital technologies mediate external visualization and interpretative mathematization. The focus is on how learners translate digital representations in pictorial or iconic form into physical-mathematical contexts. They see two main uses for digital tools. First, it can help mediate between physical phenomena and their formal representation, or between idealized models and a formal representation. For both functions, they give numerous examples for the use of different digital tools, especially video analysis tools or IR cameras, including AR and VR. In the second case above, the use of simulations, programming, or microworlds like ALGODOO is being discussed.

Thus, in this section, a wide arc is drawn from the abstract, philosophical, historical and cognitive foundations of mathematization via concrete mathematical elements to modern possibilities of visualization with digital tools. Overall, PER as a whole has deepened and broadened the view on the interplay between mathematics and physics in an educational context. Results from other fields of PER such as problem solving, representations, or digital tools have greatly contributed to the advances. Theoretical foundations, e.g., from cognitive psychology, have also supported and inspired this development. In the future, this road should be followed more intensely for being able to develop an explanatory theory from which “didactics of mathematics in physics” evolve.

CHAPTER 18

ROLE OF MATHEMATICS IN PHYSICS FROM MULTIPLE PERSPECTIVES

Gesche Pospiech and Ricardo Avelar Sotomaior Karam

Pospiech, G. and Karam, R. A. S., "Role of mathematics in physics from multiple perspectives," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 18-1–18-28.

18.1 INTRODUCTION

The relation between physics and mathematics has been discussed and wondered upon as long as humans have explored their world. One might think that the history of physics began with qualitative observations in which mathematical tools played no role. Early on, efforts were made to systematize the observations and to find underlying structures or principles enabling generalizations. Historically, among the first attempts were geometrical as well as arithmetic approaches.

The best known example are Plato's attempts to order the world in a systematic way on the basis of elementary geometric objects: regular triangles, squares and pentagons as well as regular bodies. In contrast, the Pythagoreans tried to describe phenomena in an arithmetic way with natural numbers and their ratios. A sophisticated step on the border of mathematics and physics consisted of the calculation of the volume of a sphere or a cone by exhausting the volume by thin slides, a method brought forward by Archimedes. In these approaches, mathematics served as a lens for looking at natural physical phenomena and vice versa, physical objects were described by mathematical elements. These reciprocal perspectives are characteristic of the relation between mathematics and physics. In the following steps of scientific evolution, the effort to understand and describe physical processes in detail led to the development of mathematical techniques that enabled deeper analysis. One of the most famous examples is the approach of Newton, who invented the fluxion calculus for describing motion and the changes in motion. At the same time, this example shows how entangled mathematics and physics are the mathematical description arose from physical questions and the development of the related mathematical techniques not only opened the way to physics as it is understood today

https://doi.org/10.1063/9780735425514_018

but also was at the origin of deep mathematics research. In addition, this example also shows that not only abstract thinking but also visual representations are important. The question of a notation that supports thinking and allows manipulation is by no means trivial. That a suggestive and easily manageable notation makes working with such tools much easier can be seen with the Leibniz notation of the integral. Also in recent cases, the development of a fruitful notation proved to be the key for the further progress, e.g., the Dirac notation in quantum physics or the Feynman diagrams for elementary particle physics.

Part of this fundamental interplay is that mathematics provides algorithms, formalism or structures for solving not only single problems but whole classes of problems, whereof the Lagrange formalism is an extraordinary example in that it enabled deep insights and progress. This implies that some formal approaches can be applied quite generally and can be adapted to different physical situations or phenomena. Mathematical results might carry a strong physical implication, e.g., the Hamilton formalism (classifying motions) or the principle of least action (solving very different problems) or the invariance of quantities under transformation (conservation of quantities). This endows mathematics with modeling and predicting power in physics. On the whole, mathematics provides structures for enhancing a deeper understanding of the world and especially the physics processes that take place in it.

The possible roles of mathematics in physics will be discussed in the next section. The corresponding manifold aspects should inform the teaching of physics along the educational career. From these, it can be derived:

- what students should learn and which insights they should obtain at lower secondary school, high school or university
- what should be the knowledge and awareness of (future) teachers by defining appropriate pedagogical content knowledge with respect to mathematics in physics

In the following section, we will first provide some philosophical and historical background in order to frame the discussion. To inform the discussion of the interplay of physics and mathematics from a learning perspective, we present theoretical frameworks in order to be able to describe and analyze learning processes or possible learning difficulties. An overview of the most relevant literature in this field is given in Fig. 18.1.

18.2 PERSPECTIVES ON THE INTERPLAY

The interplay of physics and mathematics has been the object of many philosophical, historical and educational considerations. Depending on the perspective of the analysis, different aspects are emphasized in each case: the potentials or chances, the difficulties or also the fascination in view of the great achievements of the interplay between physics and mathematics. Among all the scientists, theoretical physicists are taken to be the ones with the best direct insight and who therefore might

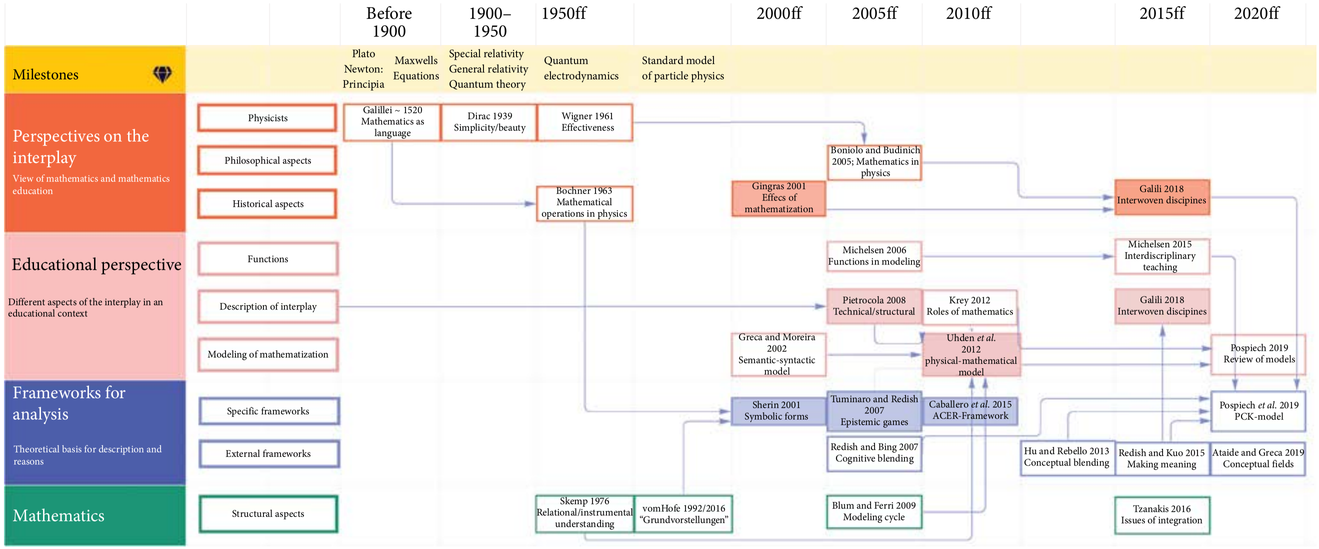


FIG. 18.1

Overview of the most relevant literature and their relation to each other.

influence the view of many others. To these belong, e.g., Dirac, not only stressing the success of mathematics as a whole but tracing it back to additional criteria, the simplicity and beauty of mathematics equations in physics as a criterion for applicability and adequateness (Dirac, 1939). An enormous influence on the thinking about the relationship goes back to Wigner, stressing the effectiveness of mathematics in physics (Wigner, 1960). Einstein focused on the complementarity of mathematics and physics with respect to exactness and world description (Einstein, 1921):

> "Insofern sich die Sätze der Mathematik auf die Wirklichkeit beziehen, sind sie nicht sicher, und insofern sie sicher sind, beziehen sie sich nicht auf die Wirklichkeit."
>
> *Translation: "Insofar as the propositions of mathematics refer to reality, they are not certain, and insofar as they are certain, they do not refer to reality."*

These different perceptions will be discussed in the following and set into the philosophical context.

18.2.1 Philosophical perspective

Since the goal of physics is to describe processes in nature precisely, the real objects have to be perceived in a way that makes this possible. This is done in the first place with the help of idealizations and physical models in the treatment of which mathematical elements of different kinds may play a crucial role. Ultimately, the use of mathematics has raised the question—and still does—why mathematics, as a body of thought seemingly independent of nature, is so well suited to describe physical processes. A fundamental treatment of this interplay from philosophical and historical perspectives would go too far here. Therefore, we restrict ourselves to pragmatic descriptions relevant to physics education. Nevertheless, we pick out particularly often cited and relevant sources highlighting different aspects. We are confident that the philosophical debate has educational value, since it is not a given that physics is intrinsically mathematical. So why mathematics is useful or effective in physics? What are its roles or in which way is it necessary for physics? Here, we present some possible answers from different philosophical positions.[1]

The world is mathematical According to this view, the world is either fundamentally mathematical, i.e., it is made of mathematical entities, or the elements that constitute the world are structured mathematically, as if they were made by a mathematician (Islami and Wiltsche, 2020, p. 159). This immediately explains why the world contains so many mathematical relations (e.g., algebraic, topological, geometrical, etc.). One of Galileo's most famous citations claiming that "The great book of nature, is written in mathematical language" epitomizes this view.

---

[1] This categorization was inspired by Boniolo *et al.* (2005) and by lecture notes produced by Professor Hans Halvorson in a Philosophy of Science course given at the University of Copenhagen.

Physics and mathematics have a common history and development A more pragmatic view is the one that recognizes the mutual influence of mathematics and physics in their historical development (Tzanakis, 2016). New problems in physics led—and lead—to new developments in mathematics, which again may lead to new physics, etc. Thus, it is no surprise that physics and mathematics is in such harmony with each other, as this fact reflects the historical development. In Kjeldsen and Lützen (2015) it was analyzed that mathematics develops most strongly when applied to solving concrete problems, not only in physics. Therefore, before the application (not only in physics), mathematical tools have to be developed and then the mathematical structures might display their own dynamics.

However, the application of mathematics in physics is not only a success story but there are also examples where mathematics at least first led to errors or misconceptions (Brush, 2015). A perhaps not so well known instructive example of how mathematics and physics interact mutually and often unexpectedly might be provided by the treatment and role of vortex theory of atoms and mathematical knot theory, which were prominent during the 19th century, then went to the background and finally had great success at the end of the 20th century up to recent developments in theoretical physics (Kragh, 2015).

Physics studies the "mathematizable" Another position claims that physics is a science that studies a small range of phenomena that can be mathematized. In this sense, mathematics puts some kind of constraint on the nature of phenomena that physics can investigate. Then, mathematics is useful for physics since physics only studies the aspects of reality that can be mathematized. Of course, these aspects change with time since new developments in mathematics can expand the range of phenomena that can be investigated by physics. This can be observed, e.g., in the case of nonlinear systems.

Mathematics is analytical priori Many view mathematics as sets of axioms and theorems derived from these axioms. Since one is free to choose the set of axioms, mathematics consists of investigating which theorems can be derived deductively from these axioms. This is like a "if …, then …"-view on mathematics: if the axioms are faithful descriptions of reality, so are the theorems derived from them. However, it is outside the domain of mathematics to judge whether the axioms are faithful descriptions of reality; this is given to the physicists. Thus, mathematics has, in principle, nothing to do with reality, but can judge the veracity of statements/theorems, given a set of axioms (Islami and Wiltsche, 2020). In other words, mathematics is useful because it gives us the rules about how to make conclusions/theorems from a set of axioms, but the latter are chosen by physicists from interactions with the world.

Mathematics as embodied cognition One last and more complex view focuses on the question "where does mathematics come from?." Supporters of the notion of embodied cognition argue that when we as humans meet the world (mainly through our senses), we (unconsciously) systematize our impressions about it after certain principles. These can be, e.g., fundamental geometric or arithmetic conceptions. Mathematics is developed either as descriptions of these principles or as a result of them, which at a very fundamental level shape our experience of reality. In this sense, mathematics itself has deep roots in the physical world, so it is no wonder that it is useful to describe it (Lakoff and Núñez, 2000).

18.2.2 Historical perspective on the interplay

In the same spirit as the previous section, we present a sketch of the historical development of the interplay between physics and mathematics. We will identify three overaching phases where the nature of this relationship is fundamentally different. This is, of course, oversimplifying a rather complex and imbricated historic process. Nevertheless, we still hope it can be useful to physics education researchers interested in reflecting on the mathematics-physics interplay.

The pedagogical value of a historical view is discussed in (Tzanakis, 1999). Using two examples (Newton's law of gravitation with Kepler's laws and the special theory of relativity) of the intertwining of mathematics and physics, Tzanakis shows that it is not the exact tracing of the historical development that is meaningful for learning, but rather that it is important to present the open questions and problems that led to the discovery or development of new concepts or ideas. The historical analysis

> "… indicates how physics may supply important examples of new and/or abstract mathematical concepts, and conversely, how simple mathematics leads to physically nontrivial results, or help to understand supposedly unintelligible physical ideas or theories." (Tzanakis, 1999)

In this respect, historical analysis can serve as a source of inspiration for a profitable lessons on pioneering ideas on the border of physics and mathematics. Most importantly, Tzanakis points out that the historical procedure can be represented as a problem-solving process, but should nevertheless be informed by the recent knowledge and insights, e.g., profit from the use of modern notation.

Pre-mathematization phase Mathematics is such an intrinsic part of physics today that we tend to think that it has always been like this. On the contrary, if we go back to antiquity, take Plato for instance, we will find a fundamental incomensurability between the i) realm of mathematics - with its perfect lines, circles, etc. that exist only in our minds and/or in the holy heavens - and ii) the social - physical world. Thus, for many thinkers at the time, one could not use mathematics' perfection to describe the imperfect objects of the real world. This incompatibility view can be found all the way until the scientific revolution, as exemplified in the following quotations [cited in Gingras (2001), our emphasis]:

> "The minute accuracy of mathematics is not to be demanded in all cases, but *only in the case of things which have no matter. Hence, its method is not that of natural science*" (Aristotle, ±350 BC).
>
> "Geometry is geometry only through the *abstract simplicity of its object*. Only that makes it certain and demonstrative. The *object of physics is much vaster*. That is what makes it difficult, uncertain and obscure. But this is essential to it: *one is not a better physicist because one is the best of geometers*" (Castel, 1743).

Mathematization phase But at some point the scientists must have rejected this incompatibility and got used to describing the physical world with mathematics. Again, this did not happen overnight, but was a long and convoluted process, where names like Galileo, Kepler, Huygens, Newton, D'Alembert, Euler, among many others, played an important role. Gingras (2001) describes this process by proposing three major effects caused by the mathematization of physics:

1. social—exclude actors from legitimately participating in discourses on natural philosophy
2. epistemological—transform the very meaning of the term "explanation" as it was used by philosophers in the 17th century; and
3. ontological—vanishing of substances such as Cartesian vortices and the luminiferous ether.

At the end of 18th-century, physics and mathematics were naturally interrelated. In fact, often the same persons, e.g., Euler, Cauchy, Gauss - would make significant contributions to both fields.

Post-mathematization phase This strong and natural interrelation does not mean, of course, that all developments in mathematics were driven/influenced by physics. The three names mentioned above are particularly famous for their works in purely abstract mathematics. But by the mid-19th century, a stronger movement of the "independence" of mathematics began to emerge, for instance, with the advent of non-Euclidean geometry and group theory. Also from an institutional perspective, "pure" mathematics begins to be legitimized as the study of structures and patterns independent of their relation to the physical world. This view that mathematics has tremendous value independent of possible applications (e.g., in physics) is passionately defended by Hardy's Apology in 1940 (Hardy, 1940). Nevertheless, even when intentionally disconnected from the "real world" in their origins, many pure/abstract mathematical concepts and theories end up being applied successfully in physics. It is in this context that we should try to understand some of the claims made by Wigner in his highly influential 1960 paper (Wigner, 1960), where he categorically states that

> "The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve."

Although it is not miraculous at all to understand why differential calculus is applicable to describe motion since it was actually created for this purpose, it is much more difficult to justify why certain properties of symmetry groups can predict the existence of fundamental particles (Wigner, 1931).

Thus, nowadays it is a widespread consensus that mathematics is essential for physics. But why exactly? What kind of "service" do mathematicians provide to physicists? This we will answer in the following section already with education in mind.

18.3 EDUCATIONAL PERSPECTIVE ON THE RELATION OF MATHEMATICS AND PHYSICS

The philosophical and historical considerations show the close and manifold relations between mathematics and physics. In particular, mathematics has become a tool of thought in modern physics that is nowadays indispensable. Therefore, the described philosophical and historical overview and the ideas about the relevance of mathematics in physics and their roles should find their expression in physics education at school as well as in teacher education, also as part of teaching the nature of physics. This justifies their importance as a topic of physics education research.

The relevance of mathematics is also reflected in the guidelines for physics education worldwide. Physics (or science) curricula such as NGSS or many others articulate as one goal of physics education the understanding of physics methods as a component of culture-based education, among them the use of mathematics and mathematical-graphical representations. If one sees these curricular requirements or guidelines, then it is a little surprising that PER has not taken up this topic for quite a while. There was some early work like (Arons, 1976; Reif and Allen, 1992; and Monk, 1994) that addressed related issues. However, the systematic investigation did not begin until around the year 2000, including most notably the work of Sherin (2001) and the work of Redish and collaborators (e.g., Redish, 2005; Bing and Redish, 2007; and Tuminaro and Redish, 2007). Furthermore, there is work on graphs, e.g., graphs of functional dependencies which are addressed in Chapter 20. Moreover, there are observed difficulties of students on all levels in mathematizing physics processes, which will be addressed with the example of physics equations in more detail in Chapter 19.

18.3.1 Teachers and teaching

Furthermore, the stance of teachers has to be addressed. Underlying every physics class is a - more or less conscious - view of the teacher of how mathematics interacts with physics to describe the world (Ataide and Greca, 2013). According to them, one can essentially distinguish three fundamentally different ways of looking at the role of mathematics in the context of physics education:

- Mathematics describes analogies between real and ideal physical objects, but is in itself disjoint from physics. This corresponds to the stance of "Mathematics is analytical a priori" described above and might tend to ascribe mathematics a more technical role as a simple tool.
- Mathematics serves as a language, where expressions must be translated and at the same time have meaning.
- Mathematics serves as an instrument that reveals structures in the world, makes them workable, and thus contributes to theory development in physics. This corresponds to the stance of "Physics and mathematics have a common development" and clearly describes a more structural role.

Each of these views influences the decision teachers are making for shaping their teaching. In a model of the pedagogical content knowledge (PCK) for teacher education concerning mathematization (Lehavi *et al.*, 2015; and Pospiech *et al.*, 2019), the aspects described in this chapter would concern the fundamental aspect "Orientation towards teaching." Corresponding views were also found in an interview study with experienced teachers (Pospiech *et al.*, 2019). Besides this fundamental aspect, additional aspects of PCK are included in the model developed on the basis of the model of Magnusson *et al.* (1999) and adapted to this specific case by Pospiech (2019).

Knowing about the described possible views seems particularly appropriate as a background for physics instruction at school or university. Such viewpoints should inform and broaden the ambivalent nevertheless often simplified image of mathematics in physics education: Mathematics is somehow inseparable from physics and on the other hand it is seen (only) as a tool. Conversely, mathematics education uses physics (only) as a reservoir of suitable examples, be it for the introduction of functions, derivatives or integrals. In order to make the interplay more visible, sometimes an interdisciplinary teaching approach is advocated (Michelsen, 2015; and Mäntylä and Poranen, 2019). These considerations are brought into the context of the discipline-culture model by Galili (2018). He describes different areas of interaction between physics and mathematics: In the central area, both are closely related to each other, be it that the tools were developed together as in the infinitesimal calculus, be it that a need in physics has triggered developments in mathematics, as can be seen in the example of the delta-function. In addition, there is another area where mathematical methods and tools can help students and teachers alike in physical analysis, such as scaling and dimensional analysis. On the other hand, deep mathematical questions concerning the nature of numbers contribute not only to the technique of approximation relevant in the context of physics idealization but also provide deep insights as, e.g., in the KAM- (Kolmogorov-Arnold-Moser)-Theorem proving the existence of invariant tori relevant, e.g., for planetary motions. Furthermore, it is emphasized that different mathematical areas such as geometric and algebraic approaches can certainly have a specific impact on teaching and learning.

In addition to this view on the interwovenness of the two disciplines, the connection between experiment and mathematics plays an important role: how does one obtain a physical law from experimental data in an inductive way? What role does mathematics play in the formulation of hypotheses, in the evaluation and accuracy of experimental results, and in their integration into the physical structure of thought? To these questions, historical examples could also provide some insights. DeBerg (1995) discusses the complexity of the connection with the example of Boyle's law: a mathematical description of nature is an approximation, and the derivation of a quantitative law from experimental data needs intuition as well as careful analysis. The theoretical backup and algebraic expressions are useful for theoretical development and clarification of concepts.

How this general interaction might mirror in education can be seen in the concrete context of the French school system, where it reads (quoted from Cisse and Dorier, 2014)

> "It would be good that […] mathematics and physics teachers from the same school support each other mutually. Physics teachers must always know at what stage of mathematics knowledge are their students and reversely mathematics teachers would gain in not ignoring some examples that they could choose, in the experimental knowledge already acquired, in order to illustrate the theories they have explained in an abstract way. (Introduction to Programmes du lycée, 1902, p. 3)"

Radtka (2015) analyzed textbooks with respect to their treatment of the relation between physics and mathematics and how this might affect its perception by the students. This might be interesting from the viewpoint that practical work (related to physics) might enhance the learning of abstract mathematical notions. All these considerations open the view onto an important aspect that could inform school curricula and lessons in physics, a deliberate distinction between a structural, a communicative and a technical dimension of mathematics for physics (Pietrocola, 2008; Uhden *et al.*, 2012; and Ataide and Greca, 2013). This stance will be clarified in the following with mathematical elements and structures that are used in physics education already at the secondary school level.

18.3.2 Roles of mathematics in physics

In this section, we attempt to describe dimensions covered and specific "roles" played by mathematics in physics. The list builds on different references and is far from being extensive. Special attention is given to aspects that have stronger educational implications.

18.3.2.1 Technical vs structural dimensions

When thinking about the role mathematics may play in physics, it can be useful to distinguish between two broad dimensions: the technical dimension and the structural dimension (Pietrocola, 2008; Uhden *et al.*, 2012; and Karam, 2014).

The *technical dimension* is associated with the instrumental character of mathematics, as described, e.g., by Skemp (1976). This is given when physicists perceive mathematics as a toolbox, where it is quite evident that it can be disconnected or detached from physics. In this case, for example, calculations are conducted without any reflection about its physical meaning, or mathematical properties are simply invoked with no physical interpretation. However, the technical dimension (with the pertaining skills) in the sense of mathematical complexity may be important in the process of problem solving (Ibrahim *et al.*, 2017).

The *structural dimension*, on the other hand, is identified when there is a clear and constant connection between mathematical elements and physical reasoning. While it may be easy to exemplify the technical dimension, the structural encompasses a greater variety of skills. They include, for example, the process of constructing a mathematical representation for a physical situation (mathematization), interpreting mathematical results physically (interpretation), using mathematics' deductive structures to derive physical theorems or make predictions of phenomena from first principles (deductive reasoning), or describing different phenomena (e.g., heat and electricity) by the same mathematical formalism (formal analogies). Analogies and different representations can be important, as shown in the example of Faraday's field lines and Maxwell's theory in (Tweney, 2010). Maxwell used these means to foster the formal derivational and calculational role of mathematics. With this, he provided a cognitive means, e.g., for thought experiments. Analogies enabled by a mathematical description in the case of electromagnetism have been analyzed in Pask (2003) or with the example of Einsteins theory of light by Gingras (2015).

Table 18.1 provides examples contrasting these two dimensions. A detailed description of this framework in the context of instruction by lecture can be found at (Karam, 2014).

Table 18.1
Technical-structural distinction concerning the role of mathematics in physics, similar to Karam (2014).

Technical (instrumental, procedural)	Structural (relational, organizational)
Blindly use an equation to solve quantitative problems (plug and chug)	Derive an equation from physical principles using logical reasoning
Focus on mechanic or algorithmic manipulations	Focus on physical interpretations or consequences
Use arguments of authority, rote memorization of equations and rules	Justify the use of specific mathematical structures to model physical phenomena
Fragmented knowledge: memorize different equations for each specific case	Structured knowledge: connect apparently different physical assumptions through logic
Identify superficial similarities between equations	Recognize profound analogies and common mathematical structures
Mathematics as a calculation tool	Mathematics as a reasoning instrument
Mathematics as just another language	Mathematics is essential to define physical concepts and to structure physical thought

In an instructive example, a physical-mathematical model of mathematization (Uhden *et al.*, 2012, see also Sec. III D) was used with regard to the technical and structural dimensions in the analysis of Planck's path of knowledge generation in the discovery of quantum physics with the blackbody radiation (Branchetti *et al.*, 2019). This analysis was then the basis of a course at the university to introduce students to the interdisciplinary argumentation and the importance of the structural role.

18.3.2.2 Specific roles of mathematics in physics

In Krey (2012, 2014), based on a survey of central philosophical positions, four specific roles of mathematics for physics with educational significance were identified:

The reduction of cognitive load Mathematics can be viewed quite generally as a representational form of abstract objects, be it, e.g., with concrete things, graphical representations or algebraic expressions. These representations help in relating the represented things to each other or in drawing conclusions about them. If the visual realization of such representations is chosen adequately, e.g., with a helpful notation - it can generally lead to "thinking made visible." In this way, cognitive load is reduced because abstract objects or complex relations no longer have to be kept in mind but are "outsourced." In addition, the representations allow for manipulations according to clear rules, which in their turn also have a cognitively relieving effect. They allow to separate the technical manipulation (technical dimension) from the interpretation and physics ladenness of the mathematical structures behind the representations.

Exactness, precision and prediction We may safely say that mathematics is the most exact science since it requires precise definitions and is based above all on the proof of statements on the basis of well-defined terms and clearly stated assumptions and relying on the rules of logic. The thoroughness of this procedure can be transferred to physics as long as the terms are defined and the assumptions are valid. This is in physics not as easily given as in mathematics, but nevertheless has led to impressive predictions of physics phenomena, among them the electromagnetic waves or antimatter. The strength of the applicability of mathematics in physics lies therefore in that the mathematical language is extremely specialized and restricted, even compared to the language in physics, and much more than the erudite language used in the classroom. The prediction of physical processes might be possible to a certain degree without formalism or mathematics, e.g., physical reasoning might predict the general behaviour of a system, such as, e.g., "the fall will get quicker and quicker." However, the precise values of the velocity or the prediction of the time of impact can only be calculated with the help of mathematics. This possibility of prediction has two aspects: one pertaining to the technical dimension, being able to use numbers and to calculate exactly. The other going far beyond calculations displaying the structural dimension (see Sec. 18.3.2.1 above): mathematics was the prerequisite of being able to develop generalized techniques for solving whole classes of problems such as the problem of brachistochrones with the Lagrange formalism, or to find physical theories such as General Relativity. However, to prove the worth of these formalisms or theories, fixed parameters (numbers) and concrete predictions with high precision were also needed to show that the given mathematical structure has physical meaning.[2] Therefore, the transfer of physical meaning to the mathematical elements and operations has to be as unambiguous as possible in order to arrive at reliable and exact predictions.

Scientific objectivity The term objectivity is generally used with different and comprises several conceptions (Reiss and Sprenger, 2020). Here we understand with objectivity the scientific objectivity in the sense that the laws of nature are independent of the subjective values or convictions and describe an objectively existing world, implying intersubjectivity (Krey, 2012).[3] This might be characterized by intuitive "faithfulness to facts" (Reiss and Sprenger, 2020). If structures in the facts are found, these can be used for explanations or prediction of phenomena. With respect to physics mathematics was extremely successful in describing or even finding such structures as mathematics can be formulated in an objective language having the same (objective) meaning across persons and cultures. Hence, as far as the laws of nature are formulated in mathematical language, objectivity is inherited by physics from mathematics. There might be a certain drawback in objectivity by the possibility of interpreting formula (and data), but this does not stem from mathematics as much as from physics and the complexity of its phenomena. But mostly the objectivity is nearly always reached in the long run by an agreement in the scientific community, abstracting from the perspective of the individual scientist (intersubjectivity) (Reiss and Sprenger, 2020). Concerning physics education, especially with respect to teaching the nature of physics, it might be fruitful to acknowledge that the finding of a law, a model or a theory in physics might depend on the context, individual values, philosophical convictions and so on, but that (reasonable) confirmation,

---

[2] In the case of General Relativity, these were given by the explanation and prediction of deviation of the appearance of stars and mercure's trajectory.

[3] In this sense also, e.g., the dependence of observation from the frame of reference is an objective fact.

appropriateness, usefulness or proving its worth of a theory is independent of the single scientist. In this sense, the scientific community plays an important role in the process of knowledge gain by discussing, critically reflecting, reasoning, or replicating results from individuals (or a working group). It might also be important to stress that even if there is in a certain sense an objective mathematical description of the world, this can mostly only be an approximation of the world because of unavoidable idealizations (Quale, 2011).

Communication Mathematics has its own signs, symbols, representations and rules for their manipulation, put simply: an own language completely different from natural language. However, the specific signs and rules in combination with the precise definitions reduce the variability or vagueness of meaning and the possibilities in applying the language. This supports focusing on essential content and reduces misinterpretations during communication. In addition, the standardization greatly facilitates the communication among people, independent of natural language and cultural background. On the other hand, the communication requires that both sides of the communication, sender and receiver, are able to encode and decode the signs and symbols and know how to manipulate them [e.g., graphical representations and algebraic expressions, see also Gingras (2001)]. With respect to physical objects, the applicability of mathematics first requires consistent modeling and joint perceptions of the physics processes, and then convincing the scientific community about the appropriateness of the model, which is a communicative act.

18.3.3 Mathematical elements with relevance in physics education

The kind of elements mathematics can provide or which have been developed inspired by physics is manifold and was developed through intricate historical processes. The span reaches from very basic to quite complex elements, sometimes mathematical structures were known before their applicability in physics became apparent, sometimes physicists used mathematics without rigorous justification, which followed later (Bochner, 1963).

In this section, we describe some often used mathematical elements from the perspective of physics education at school and university (see also Pospiech, 2008; Heck and Buuren, 2019; and Dilling and Kraus, 2022). All the addressed elements act together, but each of them highlights different aims, methods or levels of description.

Symbols and notation Symbols and notations influence the thinking about physical-mathematical problems and processes. As generally in languages also in the mathematical language suitable well-established conventions in terms of syntax, … help in communicating, describing, deducing or calculating and thus contribute to being effective in dealing with mathematical models of physics processes. So the symbols have to be characteristic (as, e.g., the Delta-function δ) and perhaps even with iconic quality (as, e.g., the Dirac-notation, which can be used with problem-adapted signs). To the main functions of a favorable notation belong simplicity (e.g., the integral sign), variability [e.g., (partial) derivatives and related symbols] and abbreviation (e.g., the index convention of relativity theory). Its objective is to facilitate manipulation, enable thinking and so on (see Sec. III B 2). The possibility of quickly writing down, easily reading and making derivations or calculations enhances communication as well as knowledge gain.

However, the best notation does not preclude the combination of the operational (technical) dimension and the structural dimension, which is even important in mathematics itself (Skemp, 1976; and Sfard, 1991). The combination with physics seems to reinforce many (students') difficulties. In a study with students in introductory mechanics, it was found that students had difficulties in simultaneously combining the meaning of the formula signs and of the symbolic equations and their manipulation (Torigoe and Gladding, 2007). Furthermore, the ability to understand symbolic equations seems central for success in physics (Torigoe and Gladding, 2011).

In the context of learning, strong conventions for formula signs such as F for force or a for acceleration, which are constantly used, facilitate communication considerably and thus reduce the cognitive load. On the other hand, the double use of signs can cause problems if, e.g., m or p may have different meanings in different contexts. This applies especially to school teaching where the students lack experience or confidence and treat different areas during a school year.

Functions and equations Physical-mathematical equations and their role in physics education as well as empirical results will be treated extensively in Chapter 19 of this book. However, we also give here some remarks.

For the description of physics processes, functional dependencies and thence functions with all the related ramifications is indispensable tools. Perhaps in this area, the strong intertwining of mathematics and physics is best visible. On the other hand, in the explicit use of functions in secondary school, some of the differences between physics and mathematics show up and cause difficulties. We have to distinguish the implicit use of functions, the explicit formulation of functional dependencies and the formal-mathematical definition and use of functions (Leinhardt *et al.*, 1990). For example, Heck and van Buuren (2019) analyze the manifold aspects of variables and functions in mathematics and physics together with their educational implications. There are attempts to reduce some of the arising problems by interdisciplinary teaching. This can be used in physics education, especially in the context of modeling (Michelsen, 2006, 2015). He proposes a horizontal as well as a vertical linking between the subjects and in the subjects respectively.

These different stages might apply to different ages of students. The fact that students are able to think in functional relations from an an early age on was demonstrated with one example taken from mathematics, related to proportionality (Blanton and Kaput, 2004). In the presented setting, it would be interesting to see if students are able to transfer their capabilities to different examples and if they become more and more fluent with its application. This example also shows that physics education could exploit the capabilities of students quite early on a qualitative-conceptual level long before more abstract and rigorous mathematical definitions are introduced. It is still open how far such a procedure could enhance physics learning in general. One example in this direction - with the keyword "visual mathematics" - was described for the case of mechanics for students of grade 9 (Mualem and Eylon, 2010). It proved to be successful for the understanding of Newtons laws and was transferred to the topic of energy by Lehavi *et al.* (2019). Also, for introducing kinematics with two-dimensional motions to students of grade 7 (age 12 to 13), such a strategy exploiting students intuitive reasoning was evaluated as highly successful (Spatz *et al.*, 2020). These examples show that qualitative reasoning within highly mathematized areas of physics is possible quite early in the school career if the basic mathematical relations are suitably visualized. At the high school or university level, the formal mathematical definition might follow. The question still to be answered thoroughly is how much students in higher grades or university might profit from the early encounter where the mathematical structures are not addressed explicitly but are the basis of the qualitative or graphical presentation. In this respect, digital tools can open up new possibilities in teaching and providing deeper insights (e.g., Erickson, 2006; and Laverty and Kortemeyer, 2012, see also Chapter 21).

Geometry When mathematization in physics is addressed, the treatment of algebra and functions or calculus are generally in the foreground. However, at the root of the mathematization of physics stood geometrical means and techniques. In this view, the domain of geometry is often underestimated in its importance for physics understanding and physics education. The advantage is that geometric properties might describe a system independent of the choice of a coordinate system and present a visual anchor. The possibility of describing geometrical objects by algebraic means, going back to Descartes, considerably enlarged the possible scope of geometry. In this sense, geometric properties or structures might be insightful in describing physical processes as exemplified in the context of kinematics and dynamics (Hestenes, 1997, 2003). Geometry also forms the based on, e.g., for general relativity or supports the description of fields in space. Overall, Hestenes (2010) recommends using the close links between geometry and (vector) algebra to enhance understanding and even proposes using a consistent mathematical formalism for physics and physics education, namely, geometric algebra.

Besides this general viewpoint, elementary geometrical elements such as points, spheres (closed surfaces), vectors, triangles, lines, or others might be of use in schools and universities. Geometry might also support mathematization because it is sometimes addressed as "experimental mathematics." In geometry, students may discover rules or theorems in an inductive way, e.g., concerning theorems on triangles. Geometrical figures help in the idealization of physics objects, in deriving proofs, or in finding relations between quantities. To be concrete, the use of geometrical methods and tools seems especially appropriate in geometrical optics. Here, important laws like Snellius law (Metz, 2014) or the Fermat Principle (Kao, 2021) can be demonstrated or proven with the help of geometry. But beyond geometrical optics also other techniques such as regarding the "area under the curve," which is used in calculus, can be seen as a geometrical description and is applied in many physics examples, mostly in kinematics. The power of such a geometrical analysis was demonstrated with the example of collisions by Theilmann (2014). So often there might be a geometrical representation of algebraic problems which might enable students to solve problems they could not solve with calculus (Ganci, 2016).

A further advantage of using geometry might lie in the more pictorial representations which are not as abstract as an algebraic expression. A drawback might lie in the fact that geometry in mathematics education is reduced in teaching compared to stochastics, algebra and calculus. This implies that students might not be sufficiently used to geometrical arguments and reasoning.

Vectors. Like many other concepts, the vector concept grew out of the interplay of mathematics and physics. Vector notation evolved for describing geometric objects in space using algebraic techniques when direction was important. This physics motivation led to the development of the concept of vector spaces in the context of mathematics (Hestenes, 2003; and Dilling and Kraus, 2022). The connection of vector spaces to the position space on the one hand and the abstractness of vector spaces on the other hand (configuration space) leads to considerable learning difficulties in mathematics and physics (Barniol and Zavala, 2014; Bollen *et al.*, 2015; and Carli *et al.*, 2020). In some aspects, such as vector subtraction, mathematics seems to be the main problem, and in other cases, such as determining a vector component algebraically, the difficulties are on the physics side. Also, the kind of representation of vectors seems to be important for learning and understanding.

Advanced mathematics. Again and again, concrete problems arising in physics incited mathematicians to make definitions more precise (definition of integral or derivative) or invent functions with specific properties (Delta-Function) or find new techniques (e.g., in most recent times describing topological insulators). Especially, the calculus of variations and the branch of functional analysis are strongly related to physical applications. Those possible advanced topics are used at universities in the context of describing physical processes and also occur in university teaching. For advanced physics and mathematics, the propagation of heat, described by partial differential equations, may serve as an example. In case of advanced mathematics, students seem to have not only physics conceptual problems but also problems in mathematical understanding beyond the difficulties of transfer (Christensen and Thompson, 2010; Kustusch *et al.*, 2014; and Bajracharya *et al.*, 2019). In the context of the description of the motion of bodies, the need for the derivative or the integration of time-dependent functions still leads to learning problems at all levels (Arons, 1976; Reif and Allen, 1992; and Basson, 2002).

Mathematical procedures. If one takes the position that the essence of mathematics is to provide theories and structures that help in solving fundamental questions, then the corresponding mathematical procedures are also of great importance to physics. Some of the procedures are

- Calculations or numerical approximations support in making predictions and studying phenomena.
- Numerical simulations help in studying complex physical-mathematical models and might help in supporting explanations or claims.
- The statistical evaluation of data helps in identifying patterns in the data and thereby finding hints to new phenomena.
- The rules of logic provide rigor to physics concerning the mathematical derivation of a theorem (physical law) with a proof.

These mathematical procedures have the status of tools, but they also foster the evolution of physics theories because they show structures or details of processes, e.g., supported by visualization, that lead to closer analysis, to the design of experiments, or to the development of models or hypotheses.

In physics, the methods of idealization and approximation can be seen as two sides of a coin (Galili, 2018). Idealization means that in the description of a physical process, parameters are neglected, which have only a small or minor influence on the process. A common example is friction in a motion or the assumption of point-like particles in an ideal gas. This physical simplification is sometimes a prerequisite for mathematical treatment at a (relatively) simple level, such as in school or university teaching. From this point of view, it is connected to a (numerical) approximation of the mathematical equations or solutions. Technical as well as structural aspects of the interplay between physics and mathematics play a role: While the decision which parameters are neglected usually has physical reasons, the numerical approximation is rather a mathematical or technical consequence.

At the core of all these procedures are the abilities of structuring and solving problems by a creative and analytical procedure, which can also be demonstrated by students (Eichenlaub and Redish, 2019). At the same time, concrete calculations should not be underestimated, especially at the school level, since they can give a feeling for a physical relation and the functional dependence of quantities.

18.3.4 Models of mathematization

The descriptions and characterizations of some concrete mathematical elements which are important in educational contexts have to be put into the context of systematically describing how mathematics is being applied or used in physics. This process of mathematization belongs to the most complex and difficult obstacles in coping with physics for most students at all levels. Mathematization plays a role in finding or deriving physical laws and in solving problems, which is in the focus of teaching and learning physics. There is a broad literature on problem solving and suitable strategies of students, but we focus here on the relation to mathematics. In order to be able to analyze strategies in problem solving, learning processes and learning difficulties of students, a detailed and flexible model of the process of mathematization is necessary. Some of these models and their similarities and differences have been discussed in detail in (Pospiech, 2019). The discussion in physics education is inspired by models of mathematical modeling from mathematics education. A simple model for physics comprises a procedure with four steps (Redish, 2017): starting from the physical system in question, its mathematical representation together with its processing towards a result and then evaluating the mathematical result in the context of the physical system. This model hides the most important thinking steps students have to do in order to arrive at a mathematical representation of a given physics problem in the first place.

On the other hand Greca and Moreira (2002) take into account the learner (or physicist) who has to build mental models of the mathematical syntax and the physics semantics and their interrelation. They describe the relation of physical and mathematical models as well as their intertwining by mental models of users or learners. Here, a physical model is defined as an idealized physical system, described in the framework of a physics theory. A mathematical model is defined by a set of mathematical statements which are syntactically correct. In their description first, the comprehension of the physics phenomenon and its translation into a physical model (the semantics) comes before the use of appropriate mathematical models (the syntax). This view has an influence on the teaching of physics, especially the use of models and modeling in order to reach an appropriate understanding of physics concepts. An often used example is acceleration. Taşar (2010) analyzed possible difficulties with acceleration concerning the interpretation of physics and mathematics by students.

A mathematical model can be seen as a set of axioms. An instructive example might be the definition of relativity theory by a set of purely mathematical axioms proposed by Andreka *et al.* and analyzed by Friend and Molinini (2016). They call a "mathematical explanation" what Greca and Moreira might have called a "mathematical model." According to Greca and Moreira (2002), a mathematical model is needed for a full description—or explanation—of a physical system (represented by a physical model). This physical model needs a translation informed by semantics into the mathematical language. This process is driven by the mental models of the user of the mathematical and the physical models. The effectiveness of this transfer process may depend on the level of experience of the user.

As we have seen, the combination of syntactic differences and semantic ladenness poses obstacles to the process of mathematization that should be analysed in detail. A model to this end was developed in Uhden *et al.* (2012). This study focuses on the procedure of increasingly mathematizing, i.e., abstracting by transferring the physical model and physical meaning into mathematical elements or structures and interpreting the arising mathematical representation back to physics in parallel. This model tries to separate the structural aspects of the interplay from the technical aspects, in this way highlighting the importance of sense making and conceptual thinking. If taken seriously, it could also be used to clarify to students the separation of technical and structural dimensions in order to support insight into the nature of physics with respect to the role of mathematics.

In view of Sec. 18.2.1, however, the preliminary stage and prerequisite of mathematization must not be forgotten: the idealization of physical reality and the creation of a physical model that only allows mathematical description. This necessary step can be seen from a philosophical perspective as one aspect of the question why mathematics fit physics: It describes what physics (or a physicist) has to do in order to be able to solve physical problems mathematically (Islami and Wiltsche 2020, p. 164). Breaking this down to the educational level, this step implies a certain mindset, namely, the ability to apply "pure" mathematics to the physical world (Islami and Wiltsche 2020, p. 167). This in turn requires the building of appropriate mental models of the physicist or the learner of physics (Greca and Moreira, 2002).

18.4 FRAMEWORKS FOR DESCRIBING THE BLENDING OF PHYSICS AND MATHEMATICS

In order to describe and analyze the process of blending physics and mathematics by students at all educational levels starting from secondary school, different theoretical frameworks have been used.

There are frameworks very specific to the interplay of mathematics and physics highlighting some characteristics. We describe

- Symbolic forms (Sherin, 2001)
- Epistemic games (Tuminaro and Redish, 2007)
- ACER framework (Caballero *et al.*, 2015)

Apart from these specific frameworks, there are theoretical frameworks borrowed from other fields of knowledge that try to explain human reasoning and thinking starting from very basic assumptions. These can therefore, beyond their original setting, be fruitful in describing and analysing the blending of physics and mathematics. Especially, those fundamental considerations broaden the view on the interplay and therefore might develop the explanatory power of why students are thinking and reasoning in the way they do. In the following section, we describe some of the most important publications first using the respective framework in physics education or being typical of it in the chronological order of their appearance in physics education research. This sequence shows very nicely how the path from a pure description of the interaction of mathematics and physics in learning physics in a narrowly defined context widens to deeper lying structures of students' thinking and reasoning in this area. This can also suggest explanations of observed student conceptions and learning difficulties, leading to reasoned suggestions for better instruction. We describe the most frequently cited frameworks:

- Conceptual blending framework (Bing and Redish, 2007)
- Framework of cognitive semantics (Redish and Kuo, 2015)
- Vergnaud's theory of conceptual fields (Greca and de Ataíde, 2019)

While the specific frameworks stem from the analysis of observed students' strategies in solving physics problems, the three external frameworks draw on general cognitive theories.

18.4.1 Specific frameworks

The core of the art of blending physics and mathematics lies in the fact that each mathematical symbol has to be interpreted with its physical meaning and in accordance with the corresponding physics theory. Therefore, the meaning of each mathematical operation has to be consistent within the physical description of a process or phenomenon.

18.4.1.1 Symbolic forms

The starting point of this framework is the idea that students' everyday conceptions and "naive physics ideas" might influence the way they understand physics equations (Sherin, 2001). This may be related to the stance of "mathematics as embodied cognition" from Sec. 18.2.1. The concrete experience of, e.g., adding or distributing something might help in grasping the abstract notions of addition or division. Such connections can be closely related to "Grundvorstellungen" described in mathematics education (vom Hofe, 1992). These show in which way mathematical operations such as addition, multiplication, or division might be interpreted with concrete everyday examples of physical conceptions (vom Hofe and Blum, 2016) in a blending of formal mathematics and conceptual knowledge.

As the choice of appropriate mathematical elements for the physical situation (see ACER- framework below) forms an important aspect of physics problem solving, it might be helpful for instruction to model students' possible understandings. Thus, a direct connection between the structures of equations and intuitive physics ideas was sought, resulting in the identification of "symbolic forms." A symbolic form is defined by a symbol template such as ¤=¤, and a conceptual schema. These are compared by Sherin to diSessa's p-prims. This viewpoint suggests that meaning does not only lie in the signs of the formula but also in the structure of the equations as such. The concept of symbolic forms has been quite appealing and influential since they were first described (see, e.g., Kuo *et al.*, 2013). They connect the technical dimension of the equation or formula with a more structural one, namely, how to represent a physics process with a mathematical operation.

Later on quite a few authors were inspired by this framework and have used it for analysis (Hull *et al.*, 2013) or further developed it (Becker and Towns, 2012; and Dreyfus *et al.*, 2017). For example, a special aspect nonetheless deserving attention is the handling and meaning making of negative numbers in physics (Brahmia *et al.*, 2020). They developed through interviews and discussions with experts the so-called NoNiP-framework (nature of negativity in introductory physics). Herein, they found three major aspects: quantitiy, relation and operation for characterizing the use of negative quantities in physics. In addition, this framework helped in finding patterns in the previously published studies.

18.4.1.2 Epistemic games

The notion of "epistemic games" was originally used to describe scientific inquiry by experts (Collins and Ferguson, 1993). Then, Tuminaro and Redish (2007) redefined this term in the resource framework and adapted it to use it for the analysis of physics problem solving with a focus on how students approach the problems. It was found that students search for solution strategies only in a quite limited space. Thus, the strategies could be described phenomenologically by finding local coherences. Six overall structures could be identified, named as "epistemic games" (Tuminaro and Redish, 2007). The six epistemic games - which are in the eyes of the authors by no means complete or finally settled – can be categorized according to more technical approaches and more structural approaches, one of them explicitly involving graphs or sketches ("pictorial analysis game"). In one game, students completely rely on the physics behind the situation without referring explicitly to mathematics ("physical mechanism game"). The concept of "epistemic games" was cited often in the context of problem solving as having a view on how students construct new knowledge. It fruitfully describes some of the strategies one can observe in problem-solving. According to its original definition, this framework has proven to be very adaptable and has been used in many ways. One example is the analysis of mathematical problems (partial derivatives) in upper division thermodynamics (Kustusch *et al.*, 2014). They were able to find again some of the epistemic games identified by Tuminaro and Redish (2007). Furthermore, they discussed that even with a pure application of mathematics, a relational understanding is necessary to arrive at meaningful expressions and to interpret them in the context of physics (see also Skemp, 1976). Furthermore, they identified epistemic games that were specific to the partial derivatives as used in thermodynamics and therefore functioned as "subgames" of the original epistemic games.

This framework has also been used in other cases in different elementary physics contexts, such as velocity change or temperature change, to analyze strategies in problem solving. In analogy, additional epistemic games have been identified, such as the analytical derivation game or "the graphical analysis game" (Bajracharya and Thompson, 2016). The interview study showed that framing (see below) played a major role in the students' ability to switch between these games and that some of the problems might lie in the interpretation of integrals or the "area under graph" (see Chapter 20). Deeper explanations were sought in the following studies. The search for new epistemic games was also helpful in identifying advanced or expert-like strategies of students in the context of mechanics (Eichenlaub and Redish, 2019). An extension towards the identification of epistemic games in problem solving at industrial work- places let to a reconsidering and expanding the notion of epistemic games with respect to specific contexts and their constraints (Hu *et al.*, 2019).

18.4.1.3 ACER-framework

Caballero *et al.* (2015) developed a framework (ACER: activation of tool, construction of model, execution of mathematics, reflection on result) based on various observations and experiences in university upper division teaching. For this purpose, they conducted a thorough analysis of different theories and approaches in physics education research on the strategies students use in solving physics-math problems. Most importantly, this framework serves to analyze and to understand students' difficulties through a common lens across numerous different aspects of mathematics in physics. Their investigations indicate that the biggest problem is not the direct application of mathematics (execution) but the choice of the appropriate mathematical tool (activation), the correct placement in the physical framework (construction) and the final examination of the result for meaningfulness (reflection). These phases remind of the model of mathematization proposed by Uhden *et al.* (2012) (or the representation of the model in Pospiech, 2019): the activation and the construction constitute the mathematization phase in that model, starting from a physical model and successively progressing in the choice of mathematical elements or structures to arrive at a mathematical model. This main phase is separated from (or only weakly connected to) the calculation phase (execution). Embedded is the interpretation circle going back from the mathematical result to the physical model and last is the validation of the results with the real situation (reflection). This framework is often used in the context of upper division physics (Ryan *et al.*, 2018; and Schermerhorn *et al.*, 2019).

18.4.2 External frameworks

The presented specific frameworks can be interpreted or organized on the basis of the following "external" frameworks.

18.4.2.1 Conceptual blending framework

In an effort to find an explanatory basis for students' strategies and problems of bringing mathematics and physics together that allows teaching and learning strategies to be developed, Bing and Redish (2007) refer to the Cognitive blending theory, developed originally by Fauconnier and Turner. This approach refers to the functioning of the human brain. Its main assumption is that "the mind combines two or more mental spaces to make sense of linguistic input in new, emergent ways." This process is called "blending." From their analyses, they infer that difficulties of students in problem solving might not be due to missing knowledge but to insufficient or inadequate blending. As observed in the "epistemic games" students seem to stick to one approach in solving a problem. It is suggested that teachers might be able to break up such a track by asking questions or giving hints that induce students to think more variedly. Therefore, this framework seems interesting for having relevance in instruction. But, as blending mostly occurs unconsciously, it might be difficult to develop suitable teaching strategies that really address the occurring difficulties. This is all the more the case as the different types of blending, namely, "single-scope blending" and "double-scope blending," are not always easy to differentiate and both can have their justification.

Hu and Rebello (2013b) used this framework to investigate the use of integrals in problem solving in introductory physics at the university. They found different ways to combine the knowledge spaces of integrals from mathematics and the physical content. Besides the possibility of connecting both spaces appropriately, there were less favorable possibilities: on the one hand, not to connect both spaces, or on the other hand, to relate the mathematics only to general physical concepts instead of to the concrete problem.

18.4.2.2 Framework of cognitive semantics

The goal of this framework is to understand how in the context of the use of mathematics as a language for physics meaning can be generated by learners (Redish and Kuo, 2015). In the context of physics education, three core ideas from cognitive semantics seem especially interesting: Embodied cognition, encyclopedic knowledge, and contextualization. Embodied cognition might be useful for developing abilities in mathematical reasoning, which seems relevant in the context of the model of "symbolic forms," described above. Encyclopedic knowledge refers to the fact that most knowledge does not have a meaning on its own but gets its meaning in a net of different concepts. With respect to the use of mathematics in physics, e.g., a mathematical equation has to get its meaning through the symbols used in it and their ladenness with meaning in the context of a physics theory. Contextualization implies that meaning depends on the context and therefore can change from situation to situation and the corresponding interpretation. Therefore, contextualization in addition builds on the enzyclopedic knowledge. Furthermore, it can be seen that students have to activate their cognitive resources. This establishes a kind of relation to the resources framework, also used in other contexts of physics education research (see Redish and Kuo, 2015 for references) or to Vergnaud's theory of conceptual fields (see next section).

As is often pointed out (see, e.g., Redish and Kuo, 2015) physicists are using mathematics in a different way and with different goals than mathematicians do and this difference is often neglected in teaching physics. An important reason might be that trained physicists implicitly use their physics knowledge in interpreting and applying the mathematical description and discard some of the mathematical subtleties as Redish & Kuo show with the example of the photoelectric effect and with a problem from electrostatics. They argue that the use and interpretation of, e.g., algebraic equations in physics depends on framing and the concrete given context (see also Chapter 19). The insights gained from the analysis based on these studies show that knowledge of mathematics alone is not sufficient. As in any language, the contexts and deeper meanings of the signs have to be accounted for.

18.4.2.3 Vergnaud's theory of conceptual fields

Greca and Ataide (2019) start from the idea of mental models that are used in the working memory of the individual when solving physics problems. This highly flexible mental model has to be complemented with more stable representations or knowledge of physics in the individual. A corresponding idea is provided by the notion of conceptual fields defined by Vergnaud originating from mathematics education. A conceptual field represents a quite broad structure stressing the relation between a "variety of situations," the competencies of the individual and possible conceptions. So, when given a certain problem, not only a single concept is activated but also related concepts, models, or theories, shortly: an entire conceptual field. By knowledge and experience, each individual forms schemata related to these conceptual fields that are used for understanding a class of situations or treating problems. Thus, schemata are activated dependent on the classification of a given situation by the individual. This is somehow related to the "contextualization" used in the framework of cognitive semantics. The observable effects of these schemata are so-called theorems-in-action or concepts-in-action (jointly forming the knowledge-in-action), which are individual invariants for treating a given situation or problem. In this framework, the goal of physics education is to change these implicit-in-actions to explicit concepts of physics as a prerequisite for enabling the students to think productively about physics. This general framework can be applied especially to connecting mathematics and physics during problem solving. Based on it, Greca and Ataide (2019) identified three quite general strategies, each related to a characteristic sequence of theorems-in-actions, namely,

1. Using mathematics as a technique tending to solve problems by trial and error. This approach might have something in common with one of the epistemic games discovered by Tuminaro and Redish (2007), namely, the "Plug 'n Chug game."
2. Trying, not always successfully, to form a link between the concepts and the mathematical structures. This could be related to epistemic games like "Mapping meaning to mathematics" or "Mapping mathematics to meaning."
3. Reasoning mathematically in a coherent way with the physics situation. Such strategies are found in different studies under the description of "Mathematics dominating physics" or "Focus on mathematics."

18.4.3 Blending of frameworks

One cannot say that any of the explanatory frameworks described above is deeper, better, or more powerful than another. Rather, each one of these illuminates a different aspect of teaching and learning mathematization in physics education. Therefore, depending on the context and the given research question, a selection and, if necessary, a combination of these frameworks might be used as a theoretical basis for research. We describe some examples of possible lines of research.

Drawing on the described frameworks in different ways Gifford and Finkelstein (2020 and 2021) looked for a tool that would allow to describe students' procedure in mathematical sense-making in the context of physics and developed the so-called categorical framework. Specifically, they categorized various cognitive moves that students make during problem solving and sense-making. The idea behind this framework is that it allows the description of small entities but not too small ones. They compare it to the "molecular size" as opposed to the "atomic size."

Hu and Rebello (2013a) start from the observation that students fail to apply their calculus knowledge from mathematics to problem solving in physics, using the example of integrals and integration in the context of electric resistance. As a theoretical background, they use the conceptual blending framework (mental space integration, see also Bing and Redish, 2007) taking into account the view of symbolic forms introduced by Sherin (2001) and the epistemic games found in Tuminaro and Redish (2007). Unlike these approaches, they look at the concepts and framing students use in combining input from several different areas leading to a new mental space, the blend. Mostly they observed unproductive blends tending to rely on mathematics and neglect the connection to the physical situation at hand. But it seems possible to influence the students' process of forming blends by inducing them to more productive paths. This approach was taken on by van den Eynde *et al.* (2020) to analyze in more detail the students' reasoning in the context of the heat equation and to find ways to visualize it. Thus, they were able to describe how students move between the different mental spaces (mathematics, physics and blended space). This approach achieves the advance of going from a static description of the blending situation to a dynamic description of the evolvement of students' solution to the given problems.

18.5 CONCLUSION AND OUTLOOK

In this chapter, we have attempted to make a broad survey of the various aspects of the interplay between mathematics and physics, especially in terms of its relevance to physics education. In doing so, it became clear which variety of views and interpretations is possible. The awareness of this variety is necessary for grasping the diversity of this interplay and its possible implications for physics education research. The task of PER would be to find a unified view over this complexity in order to be able to inform educators at university and teachers about possible problems of their respective learners and some reasons underlying them. In an outline of mathematical elements and structures used, we have tried to set a framework in which some of the difficulties can be recognized, classified, and ultimately circumvented or remedied by adequate instructional design. To this end, we have presented broader theoretical frameworks that allow us to classify and interpret the results of empirical studies. More concrete results with respect to physics equations and the core of the application of mathematics in physics will be presented in Chapter 19.

If the situation in modern physics is regarded as where mathematics is needed to define new concepts, the question arises of how the formalization of physics phenomena or processes can be prepared in education also without formal mathematical definitions and how would that influence physics learning at all levels. In addressing this question, it might be helpful to distinguish between technical skills, which need to be practiced (perhaps even separately from physics), and structural skills, which could be initiated through qualitative reasoning and connection to formal thinking. In addition, the role of mathematics as a language has to be considered.

Most empirical research in mathematization in physics is still exploratory and takes place in small-scale settings with mostly qualitative methods such as think aloud, interviews or observations. This is due to the complex and intricate processes that underly the blending of mathematics and physics in complex situations. However, there are some first assessments on handling and understanding graphs (see Chapter 20), on understanding calculus, the vector concept, or on the nature of negativity. But up to now there is no comprehensive test for assessment of how students handle or understand formulas. By now it could be worthwhile to think about the development of questionnaires which allow easier and meaningful diagnosis in class for educators and teachers. In the further course of research in this area, the goal could be an elaborated “didactics of formulas” or “didactics of mathematics in physics.”

REFERENCES

Arons, A. B., Am. J. Phys. **44**(9), 834 (1976).

Ataide, A. R. P. d. and Greca, I. M., Sci. Educ. **22**, 1405–1421 (2013).

Bajracharya, R. R. and Thompson, J. R., Phys. Rev. Phys. Educ. Res. **12**(1), 010124 (2016).

Bajracharya, R. R. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020124 (2019).

Barniol, P. and Zavala, G., Phys. Rev. Phys. Educ. Res. **10**(1), 010121 (2014).

Basson, I., Int. J. Math. Educ. Sci. Technol. **33**(5), 679–690 (2002).

Becker, N. and Towns, M., Chem. Educ. Res. Pract. **13**(3), 209–220 (2012).

Bing, T. J. and Redish, E. F., *2006 Physics Education Research Conference* (AIP Publishing, 2007), Vol. 883, pp. 26–29.

Blanton, M. and Kaput, J., *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Bergen University College, Bergen, 2004), 2, pp. 135–142.

Bochner, S., Isis **54**(2), 179–205 (1963).

Bollen, L. *et al.*, Phys. Rev. Phys. Educ. Res. **11**(2), 020129 (2015).

Boniolo, G. *et al.*, *The Role of Mathematics in Physical Sciences* (Springer, Dordrecht, 2005), pp. 5–8.

Brahmia, S. W. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010120 (2020).

Branchetti, L. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020130 (2019).

Brush, S. G., Sci. Educ. **24**(5–6), 495–513 (2015).

Caballero, M. D. *et al.*, Eur. J. Phys. **36**(6), 065004 (2015).

Carli, M. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010111 (2020).

Christensen, W. M. and Thompson, J. R., Proceedings of the 13th Annual Conference on Research in Undergraduate Mathematics Education, Mathematical Association of America (Citeseer, 2010), see https://ui.adsabs.harvard.edu/abs/2010APS..MARH42004T

Cisse, X. and Dorier, X., *Proceedings of CERME 6: Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education, Lyon, France*, Lyon Editions, edited by V. Durand-Guerrier *et al.* (INRP, 2014/2010), pp. 2682–2691.

Collins, A. and Ferguson, W., Educ. Psychol. **28**(1), 25–42 (1993).
De Berg, K. C., Sci. Educ. **4**(1), 47–64 (1995).
Dilling, F. and Kraus, S. F., *Comparison of Mathematics and Physics Education II* (Springer Spektrum, Wiesbaden, 2022), pp. 7–24.
Dirac, P. A., Proc. R. Soc. **59**, 122–129 (1939).
Dreyfus, B. W. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 020141 (2017).
Eichenlaub, M. and Redish, E. F., *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer International Publishing, Cham, 2019), pp. 127–151.
Einstein, A., *Geometrie und Erfahrung* (Springer, Berlin, 1921).
Erickson, T., Teach. Math. Appl. **25**(1), 23–32 (2006).
Friend, M. and Molinini, D., Philos. Math. **24**(2), 185–213 (2016).
Galili, I., Sci. Educ. **27**(1), 7–37 (2018).
Ganci, S., Phys. Educ. **51**(6), 065026 (2016).
Gifford, J. D. and Finkelstein, N. D., Phys. Rev. Phys. Educ. Res. **16**(2), 020121 (2020).
Gifford, J. D. and Finkelstein, N. D., Phys. Rev. Phys. Educ. Res. **17**(1), 010138 (2021).
Gingras, Y., Hist. Sci. **39**, 383–416 (2001).
Gingras, Y., Sci. Educ. **24**(5–6), 529–541 (2015).
Greca, I. M. and de Ataíde, A. R. P., *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer International Publishing, Cham, 2019), pp. 153–173.
Greca, I. M. and Moreira, M. A., Sci. Educ. **86**(1), 106–121 (2002).
Hardy, G. H., *A Mathematician's Apology* (Cambridge University Press, Cambridge, 1940).
Heck, A. and van Buuren, O., *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer International Publishing, Cham, 2019), pp. 53–74.
Hestenes, D., AIP Conf. Proc. **399**(1), 935–958 (1997).
Hestenes, D., Am. J. Phys. **71**(2), 104–121 (2003).
Hestenes, D., *Modeling Students' Mathematical Modeling Competencies: ICTMA*, edited by R. Lesh *et al.* (Springer, Boston, 2010), Vol. 13, pp. 13–41.
Hu, D. and Rebello, N. S., AIP Conf. Proc. **1513**, 186 (2013a).
Hu, D. and Rebello, N. S., Phys. Rev. Phys. Educ. Res. **9**(2), 020118 (2013b).
Hu, D. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020131 (2019).
Hull, M. M. *et al.*, Phys. Rev. Phys. Educ. Res. **9**(1), 010105 (2013).
Ibrahim, B. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 020120 (2017).
Islami, A. and Wiltsche, H. A., *Phenomenological Approaches to Physics*, edited by H. A. Wiltsche and P. Berghofer (Springer International Publishing, Cham, 2020), pp. 157–177.
Kao, W. F., Phys. Educ. **56**(4), 045014 (2021).
Karam, R., Phys. Rev. Phys. Educ. Res. **10**(1), 010119 (2014).
Kjeldsen, T. H. and Lützen, J., Sci. Educ. **24**(5–6), 543–559 (2015).
Kragh, H., Sci. Educ. **24**(5–6), 515–527 (2015).
Krey, O., *Zur Rolle der Mathematik in der Physik: Wissenschaftstheoretische Aspekte und Vorstellungen Physiklernender* (Logos, Berlin, 2012).
Krey, O., *E-book Proceedings of the ESERA 2013 Conference: Science Education Research for Evidence-Based Teaching and Coherence in Learning* (European Science Education Research Association, Nicosia, Cyprus, 2014), Vol. 2, pp. 163–169.
Kuo, E. *et al.*, Sci. Educ. **97**(1), 32–57 (2013).
Kustusch, M. B. *et al.*, Phys. Rev. Phys. Educ. Res. **10**(1), 010101 (2014).
Lakoff, G. and Núñez, R., *Where Mathematics Comes From* (Basic Books, New York, 2000), Vol. 6.
Laverty, J. and Kortemeyer, G., Am. J. Phys. **80**(8), 724 (2012).
Lehavi, Y. *et al.*, *The GIREP MPTL 2014 Conference Proceedings*, edited by C. Fazio *et al.* (Università degli Studi di, Palermo, 2015), pp. 843–853.
Lehavi, Y. *et al.*, *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer International Publishing, Cham, 2019), pp. 335–353.
Leinhardt, G. *et al.*, Rev. Educ. Res. **60**(1), 1–64 (1990).
Magnusson, S. *et al.*, *Examining Pedagogical Content Knowledge*, edited by J. Gess-Newsome and N. G. Lederman (Springer, Dordrecht, 1999), pp. 95–132.
Mäntylä, T. and Poranen, J., *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer International Publishing, Cham, 2019), pp. 247–266.
Metz, J., Phys. Teach. **52**(3), 177 (2014).
Michelsen, C., ZDM **38**(3), 269–280 (2006).
Michelsen, C., Phys. Educ. **50**(4), 489 (2015).
Monk, M., Phys. Educ. **29**, 209–211 (1994).
Mualem, R. and Eylon, B. S., J. Res. Sci. Teach. **47**(9), 1094–1115 (2010).
Pask, C., Am. J. Phys. **71**(6), 526–534 (203).
Pietrocola, M., *Connecting Research in Physics Education with Teacher Education, Vol. 2 of ICPE Book*, edited by M. Vicentini and E. Sassi (ICPE, New Delhi, 2008).
Pospiech, G., *Interdisciplinary Educational Research in Mathematics and Its Connections to the Arts Science* (Information Age Publishing, 2008), pp. 233–240.

Pospiech, G., *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer International Publishing, Cham, 2019), pp. 1–33.
Pospiech, G. *et al.*, *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer International Publishing, Cham, 2019), pp. 269–291.
Quale, A., Sci. Educ. **20**(7–8), 609–624 (2011).
Radtka, C., Sci. Educ. **24**(5–6), 725–748 (2015).
Redish, E. F., see https://arxiv.org/pdf/physics/0608268 for “Problem Solving and the Use of Math in Physics Courses” (2005).
Redish, E. F., *Key Competences in Physics Teaching and Learning* (Springer, Cham, 2017), pp. 25–40.
Redish, E. F. and Kuo, E., Sci. Educ. **24**(5–6), 561–590 (2015).
Reif, F. and Allen, S., Cognit. Instr. **9**(1), 1–44 (1992).
Reiss, J. and Sprenger, J., *The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab*, edited by E. N. Zalta (Stanford University, Stanford, 2020), see https://plato.stanford.edu/archives/win2020/entries/scientific-objectivity/.
Ryan, Q. X. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(2), 020126 (2018).
Schermerhorn, B. P. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020144 (2019).
Sfard, A., Educ. Stud. Math. **22**(1), 1–36 (1991).
Sherin, B. L., Cognit. Instr. **19**(4), 479–541 (2001).
Skemp, R. R., Math. Teach. **77**(1), 20–26 (1976).
Spatz, V. *et al.*, Eur. J. Sci. Math. Educ. **8**(2), 76–91 (2020).
Taşar, M. F., ZDM **42**(5), 469–482 (2010).
Theilmann, F., Phys. Educ. **49**(5), 537 (2014).
Torigoe, E. T. and Gladding, G. E., Am. J. Phys. **79**(1), 133 (2011).
Torigoe, E. and Gladding, G., AIP Conf. Proc. **883**(1), 153–156 (2007).
Tuminaro, J. and Redish, E., Phys. Rev. Phys. Educ. Res. **3**(2), 020101 (2007).
Tweney, R. D., Sci. Educ. **20**(7–8), 687–700 (2010).
Tzanakis, C., Int. J. Math. Educ. Sci. Technol. **30**(1), 103–118 (1999).
Tzanakis, C., *History and Pedagogy of Mathematics* (Montpellier, France, 2016).
Uhden, O. *et al.*, Sci. Educ. **21**(4), 485–506 (2012).
Van den Eynde, S. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010114 (2020).
vom Hofe, R., J. Math.-Didaktik **13**(4), 345–364 (1992).
vom Hofe, R. and Blum, W., J. Math.-Didaktik **37**(S1), 225–254 (2016).
Wigner, E., *Gruppentheorie und Ihre Anwendung auf die Quantenmechanik der Atomspektren - Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Vieweg&Sohn, Braunschweig, 1931), p. 422.
Wigner, E., Commun. Pure Appl. Math. **13**(1), 1 (1960).

CHAPTER

19

THE MEANINGS OF PHYSICS EQUATIONS IN THE CONTEXT OF THE INTERPLAY BETWEEN PHYSICS AND MATHEMATICS

Minchul Kim, Yongwook Cheong, and Jinwoong Song

Kim, M., Cheong, Y., and Song, J., “The meanings of physics equations in the context of the interplay between physics and mathematics,” in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 19-1–19-22.

19.1 INTRODUCTION

From an epistemological and ontological point of view, the mathematization of physics is highly effective (Gingras, 2001). Physics has maintained a close relationship with mathematics in terms of epistemology for 300 years and has changed from natural philosophy to today’s mathematical science. If we view the process of science and the development of scientific reasoning as epistemology, the role of mathematics is important in determining what we know and in finding out how we can decide what we know (Redish, 2017). Redish (2017) describes the role of mathematics in physics as follows:

> “In physics, maths integrates with our physics knowledge and does work for us. It lets us carry out chains of reasoning that are longer than we can do in our head, by using formal and logical reasoning represented symbolically.” (p. 5)

If we trace the historical development of physics knowledge, it can be said that physics and mathematics are highly interdependent (Gingras, 2015). There are many cases where physicists begin their research from intuitive physics principles and expand their research to the search for mathematical structures

https://doi.org/10.1063/9780735425514_019

(Zahar, 1980). Uhden *et al.* (2012) argued that Einstein emphasized the mathematical nature inherent in physics. In the process of clarifying the physics of Planck's law of radiation, Einstein used formal analogies comparing the syntactic form of equations for well-known physical systems with that for less well-understood physical systems to reveal the meaning of the latter equations. Kjeldsen and Lützen (2015) conducted a study that explored the nature of mathematics regarding mathematical concepts which have their origins in physics and emphasized the influence of physics on the historical process of defining functions.

In this chapter, we focus on how physics education treats the interrelationship between physics and mathematics. Traditional physics teaching in schools tends to view mathematics as only a tool for calculation and description. Additionally, mathematics is often seen as an obstacle to understanding the concepts of physics. In mathematics, physics is usually viewed only as a context in which abstractly defined mathematical concepts can be applied (Kjeldsen and Lützen, 2015). Students tend to list equations and imitate examples through numerical manipulation and substitution in physics learning situations and cannot easily abandon the equation-centered problem-solving strategy (Hestenes, 1987). Students who are accustomed to manipulating physics equations may still have many misconceptions about physics concepts, laws, or principles. Understanding or intuition of physics is not acquired merely by solving many problems (Knight, 2004). Arons (1997) argued that physics teaching, which involves solving stereotypical problems by simply substituting a number, contributes to a lack of critical thinking or scientific reasoning. Students' understanding of the interplay between physics and mathematics seems to be one of the most challenging teaching objectives for teachers in physics education (Pospiech *et al.*, 2019).

After all, in physics education, the diversity in the interplay between physics and mathematics is easily missed, and it is true that the mathematical aspects are often overemphasized in physics teaching. Upon reflection on the limitations of traditional teaching, it is argued that various interactions between physics and mathematics should be explicitly discussed with students in teaching and learning physics (Kragh, 2015). Lopez-Gay *et al.* (2015) said that although the dichotomy between the rigorous application of mathematics and physical intuition is prevalent, the application of mathematics helps to model physical situations.

In this chapter, we look at the interplay between physics and mathematics in physics education, mainly in terms of physics equations. Firstly, we discuss how a physics equation differs from a mathematics equation. Then, we discuss the physical meaning of physics equations in the ontological category frameworks and their epistemological role. Then, we discuss theoretical perspectives on the interplay between mathematics and physics from a modeling, blending, and epistemological belief perspective. Finally, we review empirical studies on students and teachers' comprehension of mathematization. Through this discussion, we hope to provide a more proper positioning concerning the importance and value of teaching the interplay between mathematics and physics in physics education.

19.2 PHYSICS EQUATIONS IN PHYSICS EDUCATION

19.2.1 Mathematics equations and physics equations

Physics equations are expressed with symbols representing physics concepts and include the relationship between physics concepts (Hewitt, 2011). Physics equations may serve for defining physics concepts. They model the conceptual knowledge of physics such as laws, principles, and theories using symbols as mathematical expressions.

The equations used in physics have a form very similar to those used in mathematics but have different characteristics (Kim *et al.*, 2018). First, mathematics equations express the abstract and the ideal (Lloyd and Sivin, 2002). Physics equations, on the other hand, indicate the real phenomena. In the process of modeling actual phenomena, physics equations must consider assumptions, conditions, and limitations for establishment from the beginning. For this reason, physics equations are not absolute knowledge that can be applied to all situations, whereas mathematical equations are non-situational and purely logical knowledge. In this sense, a big difference between physics and mathematical equations is that physics equations may or may not be correct depending on the context.

Second, each of the symbols in physics equations indicates a specific object, whereas that in mathematical equations does not have a specific object. That is, F, m, and a (representing force, mass, and acceleration) have abstract or concrete objects that they refer to, but symbols such as x and y included in mathematical equations represent unknown numbers and variables.

Physics equations can be said to be the result of physics research and/or the expression of physical intuition. Among Bruner's modes of representation (Bruner, 1964) about the structure of knowledge, physics equations correspond to the symbolic representation. In the history of science, scientific laws and principles in ancient times were usually expressed using geometry instead of equations. Nonetheless, some equations can be found. For example, around the 4th century, Pappus of Alexandria used a form of equations when explaining the theory of motion on an inclined plane (Dugas, 1988). Medieval interpreters who followed Aristotle's philosophy used the $v \propto F/R$ equation that speed (v) is proportional to force (F) and inversely proportional to resistance (R) (Carteron, 1975). The mathematization process using equations was strengthened with the discovery of calculus, and since Newtonian mechanics, the use of equations related to physics has increased rapidly. Important equations in celestial mechanics such as $T^2 = kR^3$ (Kepler's 3rd law) and $F = GMm/r^2$ (Newton's law of universal gravitation) also appeared around the same time (Cohen, 1956). Also, in the field of electricity, Coulomb began to treat electrical phenomena as equations by suggesting $F = kQq/r^2$, similar to the form of gravitational force. To summarize, the achievements of Kepler, Galileo, Descartes, Huygens, and Newton led to the rapid mathematization of science, and the studies of Euler, Lagrange, and Laplace made equations of scientific laws, principles, and theories similar to those in use today (Dugas, 1988).

In this way, the mode of physical expression has evolved from geometry to equations. Not all laws, principles, and definitions of physics concepts related to physical phenomena are expressed in physics equations, but many important physics ideas are expressed using these equations. This mathematization through equations contributed greatly to the development of analytical mechanics. With the advent of modern physics, the use of geometry decreased, and equations became more dominant. This can be said to be a natural result of our interest shifting from the continuum scale to the discontinuous and atomic scale with the beginning of quantum mechanics (Zalta, n.d.).

Historically, mathematization has had a positive impact on the development of physics, but this may not have been the case in physics education. For example, Sands (2019) likened the failure of mathematics to reflect physics to a broken mirror and raised the question "if the theory of thermodynamics found in textbooks and almost universally taught in universities is flawed in that the mathematical structure does not reflect the physical processes, how should the physics education community respond?" (p. 84)

19.2.2 The verbalization of physics equations

Mathematics is a language of physics and a tool for reasoning (Feynman, 1965). Pospiech and Fischer (2021) highlighted the roles of differential equations and their exponential functions with various examples such as the Laplace and Poisson equations, the Helmholtz equation, the Schrödinger equation, and Euler's equation for exponential functions. Furthermore, they categorized mathematical tools used in different schools and grade levels from numbers, units, and algebraic operations to differentiation and integration. In the integrated physical-mathematical model, Uhden *et al.* (2012) and Pospiech (2019) explained the steps of mathematization and interpretation between the world and technical mathematical operations. Geyer and Kuske-Janßen (2019) classified the different mathematical representations in physics lessons into four types: pure mathematical, verbal, pictorial, and objective representations. They also emphasized the interchange of multimodal representations, which helps in understanding physics knowledge. In particular, they presented examples to demonstrate the various possibilities for translations from a table to a graph with a focus on more technical or structural procedures. Janßen and Pospiech (2015) developed a model describing levels of verbalization of a physics equation, and Geyer and Kuske-Janßen (2019) provided an example for various levels of verbalization of a physics equation. Pospiech and Fischer (2021) presented a model developed by Kuske-Janssen for working with a physics equation and defined six levels (equation, representation as text or mathematical representation, terms of the equation, application in everyday language or specialized physics language, application in everyday language or in the language of the classroom, interpretation in everyday language of awareness) of physics equations using this model.

Physics equations are closely related to mathematics in their structure and form. In many textbooks, regarding the equation $F = ma$ expressing Newton's second law, "the net force applied to an object is the product of its mass and its acceleration (Walker, 2014)," or "the amount of acceleration is given by the acting force divided by the object's mass (Cummings *et al.*, 2004)." Halloun (2006) explained the form of equation of Newton's second law and said that the physics equation $F = ma$ is a "formal expression." For the equation $\int \vec{E} \cdot d\vec{A} = q_{enc}/\varepsilon_0$ expressing Gauss's law for the electric field, an expression that "the area for the electric field passing through the closed surface is equal to the value obtained by dividing the net charge inside the closed surface by ε_0" can also be found in textbooks (Serway and Vuille, 2012). These quite formal descriptions of physics equations are closer to a description of a mathematical structure than to a connection to a physical reality.

The use of the mathematical structure of physics equations and their connection with physical reality is also desirable in physics teaching and learning (Hansson *et al.*, 2015). The mathematical structure of a physics equation provides syntax to the physics equations to inform about the quantitative relationship of symbols, whereas the structure connected with physical reality provides semantics to the physics equations. It reveals the relationship between signs and the objects to which they are referred, enabling the physical interpretation. For example, with respect to the equation ($F = ma$), one can interpret that "an object is accelerated by a force applied to it" or "the change in the motion of an object is caused by a force." With respect to the equations of Gauss's law for an electric field, the interpretation that "the cause of the electric field is an electric charge" or "an electric field is generated around an electric charge" focuses on the physical meaning of the equations (Kim *et al.*, 2018).

These expressions have their own meanings, but it is difficult to provide students with a qualitative understanding. Nevertheless, it is claimed that descriptions about the connection with physical reality must be included (Kim *et al.*, 2018) and the meaning of a physics equation needs to be intelligible to learners beyond the formal description of the equation (Janßen and Pospiech, 2015).

Verbalization of a physics equation can help distinguish between a mathematical and a physical explanation. A mathematical explanation of an equation describes its syntax to inform about the quantitative relations of symbols, whereas a physical explanation provides semantics to the symbols and informs about the relationship between the objects to which this is applied, enabling physical analysis (Kim *et al.*, 2016a, 2018). Redish (2017) showed how concept knowledge is packed in the equation of Newton's 2nd law with the meaning of the symbol of the equation, including vector sign, equal sign, subscript, and superscript.

However, from the fact that many influential general physics textbooks present the mathematical meaning instead of the physical meaning as the definition of force or electric field, it can be inferred that the physical meaning is neglected compared to the mathematical meaning (Hewitt, 2002; and Walker, 2014).

With respect to the meaning of physics equations, their mathematical meaning has been discussed for centuries. For example, Newton in his *Principia* (translated into English by Andrew Motte in 1729, the translations revised and supplied with an historical and explanatory appendix by Florian Cajori in 1934) said, "But these propositions are to be considered as purely mathematical; and therefore, laying aside all physical considerations, I make use of a familiar way of speaking, to make myself the more easily understood by a mathematical reader" (p. 164). Although the role of mathematics in physics has grown significantly since Newton, not everyone was in favor of this strengthening of the role of mathematics. For example, Faraday was skeptical of the mathematization of electromagnetics at that time. However, the objections of Faraday and others were largely forgotten. For example, Maxwell insisted that natural philosophy should be mathematics because laws related to quantities must be dealt with in accordance with precise reasoning principles.

Due to the mathematization of physics, many people were excluded from participating in the discourse on natural philosophy (Gingras, 2001). From an epistemological standpoint, the meaning of "explanation" has changed because mathematics in mechanics is used unlikely to kinematics. Looking at ontological aspects, substances which are concrete concepts have disappeared due to the increased abstraction of phenomena (Gingras, 2015). Gingras argues that physics should provide a mechanical explanation for this abstraction and that this should not be confused with a mathematical explanation. From this, mathematical and physical explanations are interdependent. However, since Newton explained physical phenomena in mathematical language in his *Principia*, concrete physical explanation and abstract mathematical explanation have been mixed, and physics and mathematics are fused without physical explanation. Because of this, physics became no longer easy to understand. Such a fusion presently appears frequently in science education as well. If mathematical explanations of physics equations are emphasized to students, it can be a very difficult task for them to find intuitive and physical meanings of the equations. It is important to consider that difficulties in the mathematical process can cause students to give up learning physics before they have even found physical meaning.

The mathematical structure of a physics equation represents the quantitative relationship between the symbols of a physical expression. A physical expression contains many symbols, including operation symbols such as $+, -, \times, \div, =, \sqrt{}, \frac{d}{dt}, \sum, \int, \square^2, \square^3$. Sherin (2001) explained that a symbol template refers to the general structure of a mathematical expression that does not specify values or variables. In addition, there are symbols that refer to the physics terms of a constant, a parameter, or a variable. A constant is a mathematical and physical quantity that does not change. A parameter is a situational constant that does not change in a particular situation or experimental setting but changes when the situation or experimental setting changes. A variable describes physical quantities that change in equations and represents inputs and outputs in a given situation. Through the verbalization of physics equations, understanding not only the mathematical structure but also the physical meaning of physics equations is an important factor for students to learn physics.

Bagno *et al.* (2008) presented the following steps for the interpretation of physics equations:

1. Write down using physics terms the meaning of each component of the equation.
2. Show that the units on the right side of the equation are identical to the units on its left side.
3. Under which conditions can the equation be applied?
4. Describe the relationship between the components of the equation either by a graph or by a drawing.

5. Using a table analyze special / boundary cases for the equation.
6. If the equation contains assembled components, write down their physical meaning.
7. Write down, using your own words, the meaning of the equation.

19.2.3 The meanings of physics equations with ontological categories and epistemological status of scientific knowledge

Redish (2017) asserted that, in physics, mathematics codes conceptual knowledge that is not well addressed in mathematics classes, including functional dependence, packing concepts, and epistemology. Physics equations contain semantic relations among related concepts. The ontological category of concepts in physics equations has become an important theoretical issue in students' learning of physical concepts (Chi *et al.*, 1994; Arons, 1997; Etkina *et al.*, 2006; and Halloun, 2006). Several crucial semantic relations are extracted based on the distinction of ontological categories. For a correct understanding of a physical concept or theory, a correct understanding of ontological categories and an understanding of the ontological meaning of physics equations are important (Kim *et al.*, 2018). However, not all physical concepts are exclusively divided into an ontological category.

Chi *et al.* (1994) classified concepts into three ontological categories: matter, process, and mental state. Matter is something that can be contained or stored and has properties such as volume, mass, and color. Processes are viewed as procedures, events, and interactions in which the ontological properties of matter change over time. Emotional and intentional beings are categorized as mental states. At the same time, they argued that fostering students' conceptual change could require a process of recategorizing these ontological categories and emphasized that the ontological category classification is an important factor in understanding key physics concepts such as force and heat.

Arons (1997) classified the physical concepts in the dimension of ontological categories such as system, property, and interaction. A system is an object or a group of objects, such as particles, particle systems, and rigid bodies. A property is something that belongs to a system, such as mass, temperature, and density. An interaction is something that occurs between systems, such as forces and heat. Etkina *et al.* (2006) distinguished four ontological categories: object, interaction, system, and process, in the process of modeling.

The ontological categories became an important issue in science concept learning due to students' erroneous classification of the physics concepts (Chi *et al.*, 1994). Many students showed a substance-based misconception in which they tend to consider physics concepts of the process category including interactions as matter or properties of matter. Therefore, the need for a transition from the material-based perspective to the process-oriented perspective was suggested (Chi *et al.*, 1994; and Reiner *et al.*, 2000). For example, students typically showed the misconception that force and heat is a kind of property rather than an interaction. When learning the concept, it is important to view force and heat as a kind of interaction instead of a property or characteristic related to matter.

The use of ontological categories has been widely accepted as an important contribution to physics education researchers. However, recently, some details regarding Chi *et al.*'s idea that physics knowledge belongs to only one of the mutually exclusive categories were criticized. Gupta and Elby (2011) showed a shortage of this static view by presenting various concepts that are considered to be both matter and process. On the other hand, Cheong's (2016) analysis of physics concepts argued that the distinction between property and interaction could be clearer than the division between matter and process. It was also argued that the distinction between property and interaction is important in energy-related concept learning (Jewett, 2008). That is, in relation to energy, a specific form of energy is a concept corresponding to a property, and a form of energy transfer is a concept corresponding to the interaction. This issue appears to be carelessly treated in many physics textbooks. For example, the treatment of light energy as a form of energy can be easily found in many physics textbooks, although it is appropriate to view light as a form of energy transfer.

Based on the idea of an ontological category, it is possible to classify the semantic relationships between the concepts included in physics equations into two equations: state equations which "describe" state using quantities corresponding to various properties of a single system and causal equations which "explain" state through quantities that correspond to physical interaction between a system and its environment (Etkina *et al.*, 2006). While discussing the modeling of particle motion, Halloun (2006) divided physics equations into three types: state laws, interaction laws, and causal laws. Physics equations representing the state law show the change of the state descriptor that defines the state of a physical object, such as velocity, momentum, and electric current. Physics equations representing the interaction law such as that of universal gravitational force (Newton), electrostatic interaction (coulomb), and elastic binding (Hooke) are equations which involve an interaction descriptor such as a force. The physics equation representing the causal law (e.g., the work-energy principle) shows a causal relationship between the interaction descriptor (the work) representing the cause and the state descriptor (the kinetic energy) representing the effect. Similarly, Kim *et al.* (2016a) divided physics equations into those representing an interaction relationship (i.e., a quantitative relationship concerning certain interactions and relevant properties) and those representing a process relationship (i.e., the change of certain systems due to interactions) based on the ontological categories, which include system, property, and interaction. Physics equations representing the process relationship were further classified into those representing the state relationship (i.e., a quantitative relationship concerning certain properties or change of properties) which describes only patterns without explaining the origin of the pattern and those representing the causal relationship (i.e., a causal relationship between certain interactions among systems and environments and the change of properties of the system) which explains the origin of the state change (Kim *et al.*, 2016a). When classifying the meaning of physical expressions with ontological categories, Halloun (2006) uses law, Etkina *et al.* (2006) use equations, and Kim *et al.* (2016a, 2018) use relationships.

The issue of the role and status of scientific knowledge has been one of the major topics in the philosophy of science (Campell, 1920; Toulmin, 1953; and Giere, 1991, 2004). For instance, Campell (1920) distinguished between law and theory, while the former is a descriptive of observable phenomena, the latter is an inference that explains observed phenomena and regularities. These definitions is generally accepted by science education researchers concerning the nature of science (Lederman *et al.*, 2002; and McComas, 2002). On the other hand, Toulmin (1953) and Giere (2004) distinguished between law and principle. A law is a description of the regularity extracted from a phenomenon under specific conditions, whereas a principle is an idea that forms the keystone of the theory. However, conventional usage of the two terms does not always follow their definition. Concerning this mismatch, Giere (2004) pointed out that Newton's laws are called laws but are closer to principles. On the other hand, Spurgin (1984) presented physics equations by dividing them into equations expressing a definition and equations expressing a law. Kim *et al.* (2016a) defined the epistemological categories of physics equations according to the function and status of their application and categorized them into empirical law (describing relation and regularity among observable variables), definition (defining the meaning of a physics concept), and principles (the fundamental idea beyond the direct empirical verification similar to an axiom in mathematics). Overall, by synthesizing related discussions, Kim *et al.* (2016a) classified physics equations and described their various meanings of them using the ontological and epistemological categories.

19.3 PERSPECTIVES ON THE INTERPLAY BETWEEN PHYSICS AND MATHEMATICS

Physics learning is possible through the connection between physical reasoning and mathematics (Hestenes, 1997; and Redish, 2021). Technical skills in mathematics or qualitative concepts in themselves are not enough for the successful problem solving of a quantitative physics problems. Physical modeling to quantitatively explain physics phenomena is possible through interplay of physics and mathematics. Eichenlaub and Redish (2019) persuasively argued this by considering extreme-case reasoning, dimensional analysis, and estimation. Basson (2002), through the concept of acceleration, argued that physics learning requires integrating physics concepts, skills, and knowledge of mathematics. Similarly, Karam (2014) pays attention to the way mathematics is dealt with in physics instruction. He made a distinction between a technical approach involving the tool-like use of mathematics, and a structural approach concerning mathematical reasoning about the physical world. The importance of the structural and relational understanding of the interplay between physics and mathematics is also stressed by several researchers (Pietrocola, 2000; and Pospiech, 2019b). On the other hand, Kuo *et al.* (2013) pointed out that the previous view concerning conceptual reasoning required in quantitative problem-solving focus only two limited processes: selecting relevant equations and checking if a given quantitative solution is reasonable. They extend the scope of the interplay between physics and mathematics by persuasively illustrating that manipulating equations includes the blending of conceptual reasoning and formal mathematics. To express the interaction between mathematics and physical reasoning, the phrases "interplay between physics and mathematics"

(Lavagnini *et al.*, 2021), "blending of physics and mathematics" (Eichenlaub and Redish, 2019), or "mathematization" (Pospiech, 2019b) are often used in the literature. According to Pospiech (2019b), mathematization is the process of gradually transferring (with a focus on conceptual considerations) and translating (focusing on mathematics as a language of physics) physical processes or phenomena into mathematical elements and structures. However, for the term mathematization, other meanings could be inferred in addition to its previous definition. Mathematization is also used as an inference from the physical world to a mathematical model (Uhden *et al.*, 2012). Thus, mathematization could be used either in a broad sense or in a narrow sense.

When we deal with empirical studies on the teaching and learning of mathematization, there are several perspectives which could support the analysis of the related educational phenomena. Here we will briefly discuss three perspectives: mathematization as modeling, mathematization as blending, and epistemological beliefs concerning mathematization.

19.3.1 Mathematization as part of modeling

Scientific modeling is a crucial means by which scientists build their theories and knowledge (Hestenes, 1987; and Giere, 1999). A scientist's modeling and reasoning can be a very complex process, but it can be simplified for its use in educational contexts (Hestenes, 1997; and Clement, 2000). For example, Giere (2001) proposed a model for modeling that separates the real world from the domain of the theoretical model. In his diagram, the "real world" and the "data" obtained from observation or experiment are separated from the "model" representing the world, and the "predictions" are deduced or calculated from the model. On the other hand, Redish (2006) proposed a simple four-step model for mathematical modeling, which is based on the distinction between physical systems and mathematics. Uhden *et al.* (2012) proposed a model of mathematization based on the division of three domains: the real world, physical systems, and mathematics. This model seems to reflect both the division of the real world and the theoretical domain and the division of the physical world and mathematics in the theoretical domain. Considering the complexity of physical and mathematical modeling, such a distinction between three domains could be an important cornerstone for the analysis of theoretical and empirical studies.

Physical reasoning, including quantification, is achieved through complex interactions and shifts among the three domains (Uhden *et al.*, 2012; and Czocher, 2018). Since these processes are very complex, it cannot be expected that a single model would sufficiently describe all the complex processes. Instead, several models have been proposed, each of which focuses on specific aspects of these processes (Pospiech, 2019b). For example, Czocher (2018)'s model focused on an evaluation process that judges how well the transition from the real world to mathematics has progressed. A dominant mathematical model by Blum and Ferri (2009) is divided into seven stages, including "constructing," "simplifying & structuring," and "mathematizing," "working mathematically," "interpreting," "validating," "exposing," each of which may require very complex reasoning. For example, the "simplifying & structuring" stage in their model is related to idealization, which is one of the crucial aspects in modeling, and the process of idealization may require very complex thinking (Niaz, 1999; Song *et al.*, 2000; and Forjan and Slisko, 2014). Because of this complexity, the focus of empirical research concerning physical and mathematical modeling could be diverse.

Modeling is not a process that could be achieved merely by utilizing generalized procedural knowledge unrelated to content knowledge (Lehrer and Schauble, 2006). Rather, the fact that content knowledge and process knowledge could be integrated during modeling is considered a crucial advantage in science teaching and learning through modeling (Halloun, 2007). Therefore, by its nature, conceptual knowledge plays an important role in the process of physical and mathematical modeling. This point will be discussed in more detail in the following framework of mathematization as blending.

19.3.2 Mathematization as blending

Sherin (2001) presented a method for in-depth analysis on the way in which conceptual knowledge is applied in physical-mathematical modeling by introducing the notion of symbolic form. The symbolic form has two components: a symbol template, which is a simple schematic for a mathematical form, and a conceptual schema, which is an element of conceptual knowledge concerning physics. For example, by combining the conceptual schema of "balancing of competing influence" and the symbol template of "□=□," which in mathematics means the equality of two terms, it is possible to generate an equation representing an equilibrium of two forces of gravity and air drag in an object falling with terminal velocity. The symbolic form has characteristics similar to the phenomenological primitives proposed by DiSessa (1993) in the context of conceptual learning research. Sherin's view of the blending of conceptual knowledge and mathematical form has been further developed through various studies (Hu and Rebello, 2013; Kuo *et al.*, 2013; and Eichenlaub and Redish, 2019). However, students taking a physics course do not always think of physics equations as modeling and representing the physical world. Therefore, blending is very challenging for them (Pospiech and Fischer, 2021b).

In doing so, the "conceptual blending" perspective of Fauconnier and Turner (2008) has become a crucial theoretical framework for analyzing the blending of mathematics and physics during the problem-solving process. Conceptual blending is suggested as an alternative framework to the conceptual metaphor proposed by Lakoff and Johnson (1980), an influential theory on cognitive linguistics (Evans, 2006). Conceptual metaphor seems to be influential in the development of physics concepts and the learning of these concepts (Brookes and Etkina, 2007; Amin *et al.*, 2015; and Close and Scherr, 2015). The conceptual metaphor perspective also provided a persuasive argument that reasoning through metaphors is a crucial tool in reasoning in mathematics and mathematics learning (Lakoff and Nunez, 2000; and Danesi, 2007). For example, a function could be treated as a vector by representing the function as a linear combination of basic functions. Here, we can find a conceptual metaphor that a function can be seen as a vector. A framework of the metaphors or analogies clearly divides a source domain and a target domain, and the transfer in reasoning should be unidirectional from the source domain to the target domain (Gentner, 1983; and Duit, 1991). The conceptual blending perspective differs from this framework in that a blended space is constructed from several input spaces, and a new emergent structure that was not present in any input space is created (Fauconnier, 2001). From this point of view, physical-mathematical modeling is neither confined to the mathematical domain nor to the physical domain. This kind of modeling is not a unidirectional process of transfer from one domain to the another. Rather, inference proceeds in a more integrated way in a blended space where mathematics and physics are mixed. Blending could appear in various dimensions: a hybrid of physics and mathematics (Kuo *et al.*, 2013), a hybrid of various ways of representation (Gire and Price, 2014), a blending of mathematics (Zandieh *et al.*, 2014), and a hybrid of conceptual knowledge (Hrepic *et al.*, 2010). Considering this diversity, the conceptual blending perspective could be a powerful theoretical tool to analyze the complex interaction between physics and mathematics in solving problems.

19.3.3 Epistemological beliefs concerning mathematization

When students solve quantitative physics problems in addition to their physical conception and mathematical knowledge, their beliefs about the nature of knowledge and learning, i.e., epistemological beliefs, could significantly influence their behavior (Hammer, 1994; and Hammer and Elby, 2002). Students who regard the knowledge of physics as a collection of facts and equations and view it as disconnected from everyday thinking may conceptualize physics learning as memorization and may have difficulties in deep learning. Conversely, if learners view physics as a coherent system of knowledge and understand mathematical forms of physics as a way to express physical ideas, they will be more likely to become successful learners. Like the learner's conception, it is possible that a learner's epistemological beliefs might not be general beliefs that are independent of the situation (Mortimer, 1995; Hammer, 2000; and Hammer and Elby, 2003). Instead, they can be activated according to the situation and given problem. Students' epistemological beliefs may be different in accordance with contexts and sometimes may even be conflicting. Teachers should guide learners so that they can activate appropriate ideas from their epistemological resources for successful teaching and learning (Elby, 2001).

Ideas activated from the learner's epistemological resources could be related to the student's perception of the problem as well as to the type and nature of the problem. That is, the student's perception of "what is going on here?" for a given task or problem is crucial (Hammer and Elby, 2003). In this way, the recognition and judgement of students about the type of knowledge that is appropriate to apply to a specific situation is called epistemological framing (Bing and Redish, 2009a; and Scherr and Hammer, 2009). A person's framing of a situation influences his or her judgment about what is relevant to the situation and what should be ignored. In this respect, students' epistemological beliefs about physics learning and quantitative physics problem solving, their epistemological resources, and their epistemological framing of situations are all important in learning and problem solving because these issues can influence students' behavioral patterns in problem-solving situations. Tuminaro and Redish (2007) extracted six typographical scripts from students' problem-solving patterns, referred to as epistemic games. The list of epistemic games includes "Recursive Plug-and-Chug," "Mapping

Meaning to Mathematics," "Mapping Mathematics to Meaning," "Physical Mechanism," "Pictorial Analysis," and "Transliteration to Mathematics." Activated epistemic games from the six types could vary depending on the maturity of the learner and the type of problem. These epistemological beliefs and related behaviors are critical issues in the study of the interplay between physics and mathematics (Bodin, 2012; Greca and de Ataíde, 2017; and Modir *et al.*, 2017).

19.4 STUDENTS AND TEACHERS COMPREHENSION OF MATHEMATIZATION

19.4.1 Students' comprehension of mathematization

Students have demonstrated various difficulties concerning the meaning of physics equations and their utilization. For instance, Redish and Kuo (2015) described the process of students misinterpreting a physical situation by finding physics equations related to the problem first and then interpreting the physics equations only mathematically and grammatically. This process demonstrates students' misunderstanding of mathematization as many students tend to consider only mathematical relationships when interpreting and applying physics equations. It is crucial for students taking a physics course to understand a wide range of perspectives on mathematization related to the verbalization of physics equations, the meanings of physics equations with ontological categories and epistemological status of scientific knowledge, and the interplay between physics and mathematics.

However, we can easily find students who have trouble describing the meaning of physics equations (Bagno *et al.*, 2008). Uhden (2016) presents five areas of student difficulty in solving physics problems related to the interaction between physics and mathematics. These five areas are (1) difficulties in understanding physics functions and equations, (2) schematic-technical approach, (3) interferences with the students' everyday experience, (4) basic physical or mathematical difficulties, and (5) the confusion of concept definition vs concept image. Geyer and Kuske-Janßen (2019) showed students' difficulties with mathematical representations in physics. Redish and Kuo (2015) analyzed that the difficulties might be due to the fact that mathematicians and physicists load meaning onto symbols differently. Redish (2017) compared two physics majors solving problems related to electromagnetics and emphasized that it is more important to interpret the physical situation and apply the physics equations to solve the problem than to first select physics equations suitable for solving the problem. In addition, in his research, it was shown that the students had more difficulties in understanding the physics equations than the mathematical ones because physics equations have a bigger number of symbols and fixed conventions than mathematical ones.

Heck and Buuren (2019) showed the different ways of using variables, equations, formulae, functions, and operators such as the equal signs in mathematics and physics when they explained the students' understanding of algebraic concepts. Students often confuse the meaning of the symbols in physics equations. In particular, when a number is substituted for each symbol in a physics equation, the fundamental meaning of “variable” is confused (Mach, 1989). For example, in 1 m = 100 cm, m is not a variable but a unit. Besides students, it has been found that even professors and high school teachers demonstrate difficulties with these kinds of physics concepts (Lochdead, 1981). It is thus important to teach the symbols and meaning of expressions to students in basic physics courses (Hewitt, 2011).

Kim *et al.* (2016b) found in a study on science-gifted high school students’ perceptions of physics expressions that students perceive physics equations related to principles or definitions as more important than equations related to empirical laws. They also experience more difficulties related to physics than to mathematics when applying physics equations. With respect to mathematical difficulties, it was revealed that most of them are related to the setting of signs.

In a study on students’ understanding of the ontological and epistemological meaning of physics equations, Kim *et al.* (2020) showed that students had difficulties in understanding and classifying interaction and causal equations rather than state equations. In addition, the importance of the classification according to the framework of ontological categories was emphasized in the context of understanding the meaning of physics equations. Furthermore, this study emphasized the need to learn the epistemological and ontological meanings of physics equations because students tend to understand physics equations just as empirical rules, not as definitions or principles.

Empirical studies concerning students in secondary school showed different students’ difficulties and misunderstandings of mathematization. For instance, Uhden and Pospiech (2009) studied the problem-solving strategies of 15–16-year-old students and looked at their difficulties in using units. It was found that students focused on the instrumental role when using mathematics in physics problems. In a longitudinal study by Roorda *et al.* (2015) of 10 students over the course of two years (from grades 10 to 12), it was confirmed that it took students time to acquire a single mathematical procedure, and much more time was needed to form a broad and connected understanding of it. Meli *et al.* (2016) observed the problem-solving process of first-year high school students in solving physics problems that require blending with mathematics. Instead of starting with a physics idea for the problem, the students showed that they directly utilize the mathematical form, and this bias sometimes causes them to have difficulty in solving problems.

Despite the difficulties observed in secondary school students, some positive findings have also been reported. Malone (2008) compared the effects of model-based and traditional teaching on problem-solving at the high school level. It was concluded that a model-based instruction had an affirmative effect on students’ performance on the FCI test and problem solving and that the students’ knowledge structures in model-based instruction showed more expert-like procedures than in traditional instruction in those students tend to pay attention to the surface-like features of the given problems. Pospiech and Oese (2014) investigated the views of eighth grade students about the role of mathematics. Students showed both positive and negative thoughts about the role of mathematics in physics learning. Furthermore, concerning the role of mathematics, they showed a structural view at an elementary level, such that mathematics might help with prediction in addition to its technical role.

Empirical studies at the university level have been performed more intensively than studies focused on the secondary school level. Actually, theoretical frameworks for mathematization, discussed in previous sections, have been developed mainly in the context of studies at the university level. University students also showed significant difficulty and misunderstanding if the issue is explicitly taught in the classroom. For instance, Rebello and Cui (2008) investigated the transfer of learning from mathematics to physics in problem solving in a university physics course. They demonstrated that the transfer of learning from mathematics (structured domains) to physics (relatively unstructured domains) should be examined from multiple perspectives of transfer. They also argued that the main difficulty for students in their study does not lie in their failing comprehension in mathematics. Instead, students showed inability to apply mathematics to physics problems. Planinic *et al.* (2013) investigated university students' understanding of graphs in different domains: mathematics and physics (kinematics). Their results suggest that mathematics without context is easier for students than solving physics problems which require an understanding of context. Niss (2017) analyzed the obstacles that college students face in the process of problem solving and extracted several patterns such as lack of mathematical object definition. Mason and Singh (2016) developed a questionnaire to investigate attitudes and approaches to problem solving and compared the perceptions of introductory physics students, Ph.D. students, and physics experts. They found that introductory students were in general less expert-like than the other groups.

These studies concerning students' understanding and beliefs suggest ways of instructional improvement. Several researchers have mentioned that it is important to understand the physical meaning along with the mathematical structure of physics equations in order to understand their meaning clearly. Arons (1997) argued that although many students learn physics through solving problems from problem-banking, there are few opportunities for higher-order thinking or reasoning about physics. Sherin (2001) pointed out that important and diverse meanings of physics equations were missed because students tended to regard them just as tools for solving problems.

Knight (2004) also argued that students may have misconceptions about physics concepts or principles even if they become accustomed to calculations in the process of solving many problems using physics equations. He also mentioned the necessity of harmonizing learning at the procedural level related to problem solving. Uhden *et al.* (2012) argued that in order to overcome this issue, physics learning should start with qualitative concept learning and then move to quantitative computational learning. Kneubil and Robilotta (2015) claimed that asking questions such as "What is the relationship between variables in a physical expression?" or "Why does a physical expression have a specific structure?" is crucial in order to apply mathematics in physics. Redish and Kuo (2015) argued that the reason students with high academic achievement in mathematics had difficulties when using mathematics in physics was due to the difference in the language used in physics and mathematics. They found that most of the students who participated in the study interpreted the equation $E = F/q$ as mathematical grammar representing a proportional relationship, ignoring the physics situation that the electric field is independent of the test charge. To remedy this issue, the study argued for the necessity of teaching and learning in a way that blends the physical meaning and mathematical structure of physics equations. Karam and Krey (2015) suggested that learning related to physics requires a shift from a simple calculation and description through numerical substitution to explanation and understanding physical meanings. Bing and Redish (2009) analyzed students' problem-solving processes using mathematics in the context of upper-level undergraduate physics courses. They argued that the notion of epistemological framing, which consists of calculation, physical mapping, invoking authority, and math consistency, could reveal failures in communication and could help a teacher better understand the cause of students' difficulties.

Empirical research may lead to more specific suggestions. For instance, Bagno *et al.* (2008) suggested a classroom activity focused on the interpretation of equations, described above in ll.2.2, to overcome students' limited understanding of basic equations in physics. Their activity guides the students to clearly describe the components of the equation, relationships between the components, and the conditions of applicability. This intervention enhances the students' understanding of the equation. Bodin and Winberg (2012) investigated the role of beliefs and emotions in quantitative problem-solving at the university level. While students' conceptual knowledge and expert-like belief were important factors in predicting student performance, students' evaluation of "the usefulness of physics in daily life" had little effect on their achievement. This suggests that instruction should focus on the development of students' epistemological beliefs. Greca and Ataíde (2017) also found that epistemological aspects of the role of mathematics in physics seem to be a significant factor for students to succeed in learning.

Integrated approaches between physics and mathematics have also been suggested. For instance, Dunn and Barbanel (2000) suggested a model for an integrated course concerning electricity and magnetism in physics and vector calculus in mathematics courses. As substantial differences in context and notation could hinder the transfer of knowledge between mathematics and physics, they advocated that the integration of mathematics and physics could reduce students' difficulty. Michelsen (2006) pointed out that in the traditional teaching of mathematics, it is difficult for students to transfer knowledge learned in mathematics to a new situation in science. He suggested an alternative method of teaching the concept of function based on modeling activities in an interdisciplinary context between mathematics and science. In a later study, Michelsen (2015) also suggested several exam projects with the goal that student-teachers are able to articulate a broad spectrum of interdisciplinary themes capturing the relationship between mathematics and physics. Jensen *et al.* (2017) argued that mathematization is not normally dealt with in mathematics-related subjects, and that physics could be a good candidate to deal with mathematization. They presented examples of unformalized problems that are solvable by means of mathematization in the context of physics.

The convergence of physics and mathematics could be beneficial to mathematics education. Marrongelle (2004) investigated how university students' knowledge of physics affects solving mathematical problems in an integrated class between physics and mathematics. He contrasted two types of students which are contextualizer and language-mixer, when using knowledge of physics to solve mathematical problems. Carrejo and Marshall (2007) utilized the context of kinematics to illustrate the meaning and role of abstraction in mathematical modeling. Thus, physics contexts could be utilized in the context of mathematics education.

19.4.2 Teachers' comprehension of mathematization

Pedagogical content knowledge (PCK) refers to knowledge that makes a concept comprehensible to students (Shulman, 1986). Park and Oliver (2008) viewed the PCK as a complex construct consisting of six elements. They emphasized the point that "PCK was developed through reflection-in-action and reflection-on-action within given instructional context," and that students, especially with regard to their misconceptions, had an important impact on PCK development. This implies that the relation between reflection and PCK development should be investigated.

PCK is often described as the competency to transform several types of knowledge for teaching, including subject matter knowledge (Magnusson *et al.*, 1999). In line with this point, PCK in the context of mathematization should include knowledge of various representations such as formal equations, graphs, verbal language, and transformation among the representations (Etkina, 2010). It is a complex construct which could relate to various factors such as students' (mis)conceptions or difficulties, instructional strategies such as utilizing various representations, cognitive demand required to complete math tasks, knowledge of curriculum and media, etc. (Krauss *et al.*, 2008). Baumert and Kunter (2013) also analyzed the multidimensional characteristics of PCK in the context of mathematics teaching and pointed out that it is significant to pay attention to this multidimensionality and the interplay between cognitive and motivational dimensions.

In addition, pre-service and in-service teacher education programs should provide opportunities to comprehend students' conceptions, understandings, and reasoning types. In the context of physics education, Etkina (2010) described the pedagogical practices used in the Rutgers Physics/Physical Science Teacher Preparation program. Kirschner *et al.* (2016) designed an inspection tool to examine the professional knowledge of physics teachers in the dimensions of pedagogical knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK).

Concerning teachers' pedagogical practices of mathematization, teachers' limited understanding of the subject could be problematic. For instance, Mulhall and Gunstone (2008) investigated teachers' views about the role of mathematics in physics. They found that a teacher of traditional teaching as well as a teacher utilizing the conceptual change approach did not appear to play sufficient thought of the role of mathematics in physics. They concluded that these issues should be explicitly taught and dealt with in the teacher education program. Siswono *et al.* (2017) discussed secondary teachers' mathematics-related beliefs and knowledge of mathematical problem-solving in Indonesia. They concluded that teachers did not show a high consistency in their mathematics-related beliefs, and teachers' beliefs tended to have a close relationship with teachers' limited knowledge about problem solving.

Teachers' expert-like understanding of mathematization itself does not guarantee effective teaching on the subject. For instance, Freitas *et al.* (2004) showed that teachers' conception of physics problem-solving could be incoherent to their practice. Thus, the reflection on the gap between conception and practice is required. Başkan *et al.* (2010) investigated the views of physics and mathematics teachers concerning the integration of the two disciplines. They found that teachers believed that integration between the two disciplines is necessary, but they didn't have clear ideas about the way of integration. Lehavi *et al.* (2017) analyzed physics teachers' classroom activities at the middle school level. They concluded that the teachers showed little awareness of possible student difficulty concerning the interplay of physics and mathematics. These studies suggest that special teacher training programs that explicitly deal with the interplay of physics and mathematics are required.

Many studies have been conducted to improve the understanding of the interaction between physics and mathematics in the context of physics teaching and learning. One of the ways to support physics learning concerning mathematization is an interdisciplinary approach to teaching mathematics and physics (Pospiech and Fischer, 2021a). The findings of a program which was launched in Austria to promote integrated teaching in mathematics and physics after TIMSS in 2000 were that teachers could better gain insight into students' difficulties through the program and that students were able to understand the physics concepts more fully as a result. On the other hand, Lehavi *et al.* (2017) found that teachers' PCK related to the interplay between physics and mathematics fits the PCK model of Magnusson *et al.* (1999). Etkina *et al.* (2018) mentioned teaching tasks and Student Energy Targets (SETs), which represent a final set of understanding that teachers would like all students to build through their energy learning experiences. With these tasks and targets, the importance of reference points, the distinction between vectors and scalars, the meaning of positive or negative signs, the connection of mathematical and physical meanings, the application of the meaning of physical conservation to the meaning of mathematical invariants, and the choice of system and analyzing energy in relation to the process as an ontological category were emphasized. In addition, the necessity of the use of appropriate verbal, mathematical, and graphical/pictorial representations and their interpretation were also emphasized. Adorno *et al.* (2019) showed that physics equations are used in the explanatory stage of a 5E-based learning environment (Engagement – Exploration – Explanation – Elaboration/Extension – Evaluation) in a workshop on various aspects of the Hall Effect.

Pospiech *et al.* (2019), in their study on the views and strategies of teachers related to the role of mathematics and physics in physics classes, described a basic structure of the PCK model. In this study, they analyzed teacher conceptions of the interplay between physics and mathematics. All teachers in this study held the view that mathematics is important in physics because of the daily routine of teachers and the requirements of the school curriculum. In particular, it was possible to determine teacher tendency to recognize the importance by dividing the technical and structural roles of mathematics.

Concerning teacher training, several researchers pay attention to technology-enhanced PCK (TPCK) which is the integration of the knowledge of subject matter, the knowledge of technology, and the knowledge of teaching and learning (Niess, 2005). Vogel *et al.* (2007) utilized a multimedia learning environment to enhance students' supplantation of mental operations on graphs. BouJaoude and Jurdak (2010) utilized microcomputer-based labs (MBL) to promote mathematization and the integration of physics and math.

19.5 CONCLUSION

The broad discussion of the differences between physics equations and mathematics equations as well as discussion of the physical meaning of physics equations including both ontological and epistemological aspects emphasizes the importance of proper conceptual understanding of physics equations.

We have also discussed three perspectives on the interplay between physics and mathematics: the perspective of modeling, blending, and epistemological beliefs. The perspective of modeling for mathematization reveals the complexity of physical-mathematical modeling in detail. The perspective of blending could be useful to analyze dynamic processes related to conceptual reasoning, including the blending of two domains (Kuo *et al.*, 2013). The perspective of epistemological beliefs seems to be a crucial factor to account for students' performance when the problem solving. Thus, in curriculum reform concerning mathematization, teachers should critically consider the issue of students' epistemological beliefs. Students' understanding, difficulties, or views of the issue clearly show that an explicit approach, especially regarding their epistemological beliefs, is essential to improve their competency. Studies concerning teachers pointed to a difficulty in the reform of school practice on this issue. Teacher preparation programs should enhance their understanding of the issue and encourage them to make pedagogical changes in the classroom.

Considering the complexity of mathematization and related educational issues, the development of effective teacher preparation programs could be a highly challenging task. Redish (2017) believes that effective learning and teaching methods to use mathematics in physics teaching have not been sufficiently investigated, despite many studies which have theoretically analyzed the issue and revealed students' understanding, difficulties, and views on the topic. As discussed by Pospiech (2019), the next step would be to develop effective materials and programs and evaluate the effectiveness of these reforms.

Physics and mathematics have historically greatly influenced each other and contributed to knowledge development in each field. In physics, mathematics is fused with physics knowledge, and physics is developed through formal and logical reasoning about symbols. It has been argued that this interaction should sometimes be mentioned in the field of physics education (Kragh, 2015). However, despite opinions on the interdependence of physics and mathematics and the need for its explicit education, physics education in schools often regards mathematics only as a simple tool for calculation. Sometimes, mathematics is even regarded as an obstacle to understanding physics concepts. In addition, despite many studies on physics learning and teaching methods, there are relatively few studies and teaching methods on how to effectively use mathematics in physics (Redish, 2017).

In physics and physics education, the interaction between physics and mathematics expresses itself through physics equations. Historically, the representation of physics knowledge has evolved from geometry to equations. Physics equations have a form similar to mathematical equations but have a variety of properties. According to research on their form and meaning, physics equations are divided into mathematical forms and physical meanings. In addition, according to a study on the physical meaning of physics equations, mathematization of physics from an epistemological and ontological point of view is effective in learning physics (Gingras, 2001). In this regard, a study on the meaning of various physics concepts included in physics equations according to the classification of ontological categories was also conducted. Thus, it was possible to understand the various physical meanings of physics equations beyond the mathematical-grammatical interpretation of physics equations. Through understanding the relationship between each concept included in a physics equation and the status of knowledge that a given physics equation represents (such as definitions, laws, principles, etc.), effective learning of physics knowledge can be expected.

In the study of teachers' perspectives on mathematization, most physics teachers recognized the importance of the role of mathematics in physics. In addition, physics teachers felt familiar and comfortable starting classes by solving physics problems using familiar physics equations.

The importance of research on the interaction between physics and mathematics, including physics equations, has been consistently recognized. Through these studies, valuable hints for improving students' understanding of physics knowledge and finding effective ways of teaching and learning physics for preservice teachers can be derived.

REFERENCES

Adorno, D. *et al.*, *Concepts, Strategies and Models to Enhance Physics Teaching and Learning* (Springer, 2019), pp. 61–71.
Amin, T. G. *et al.*, Int. J. Sci. Educ. **37**(5–6), 745–758 (2015).
Arons, A. B., *Teaching Introductory Physics* (Wiley, 1997).
Bagno, E. *et al.*, Phys. Educ. **43**(1), 75–82 (2008).
Başkan, Z. *et al.*, Proc. Soc. Behav. Sci. **2**(2), 1558–1562 (2010).
Basson, I., Int. J. Math. Educ. Sci. Technol. **33**(5), 679–690 (2002).
Baumert, J. and Kunter, M., *Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers: Results From the COACTIV Project* (Springer, 2013), pp. 25–48.
Bing, T. J. and Redish, E. F., Phys. Rev. Spec. Top. Phys. Educ. Res. **5**(2), 1–15 (2009).
Blum, W. and Ferri, R. B., J. Math. Model. Appl. **1**(1), 45–58 (2009).
Bodin, M., Phys. Rev. Spec. Top. Phys. Educ. Res. **8**(1), 1–14 (2012).
Bodin, M. and Winberg, M., Phys. Rev. Spec. Top. Phys. Educ. Res. **8**(1), 1–14 (2012).
BouJaoude, S. B. and Jurdak, M. E., Int. J. Sci. Math. Educ. **8**(6), 1019–1047 (2010).
Brookes, D. T. and Etkina, E., Phys. Rev. Spec. Top.: Phys. Educ. Res. **3**(1), 1–16 (2007).
Bruner, J. S., Am. Psychol. **19**(1), 1–15 (1964).
Campell, N. R., *Physics the Elements* (General Books, 1920).
Carrejo, D. J. and Marshall, J., Math. Educ. Res. J. **19**(1), 45–76 (2007).
Carteron, H., *Articles on Aristotle 1. Science* (Duckworth, 1975).
Cheong, Y. W., Sci. Educ. **25**(5–6), 611–628 (2016).
Chi, M. *et al.*, Learn. Instr. **4**, 27–43 (1994).
Clement, J., Int. J. Sci. Educ. **22**(9), 1041–1053 (2000).
Close, H. G. and Scherr, R. E., Int. J. Sci. Educ. **37**(5–6), 839–866 (2015).
Cohen, I. B. *et al.*, *Franklin and Newton: An Inquiry Into Speculative Newtonian Experimental Science and Franklin's Work in Electricity as an Example Thereof* (American Philosophical Society, 1956).

Cummings, K. *et al.*, *Understanding Physics* (Wiley, 2004).
Czocher, J., Educ. Stud. Math. **99**(2), 137–159 (2018).
Danesi, M., Stud. Philos. Educ. **26**(3), 225–236 (2007).
DiSessa, A., Cognit. Instr. **10**(2–3), 105–225 (1993).
Dugas, R., *A History of Mechanics* (Dover Publication, Inc, 1988).
Duit, R., Sci. Educ. **75**(6), 649–672 (1991).
Dunn, J. W. and Barbanel, J., Am. J. Phys. **68**(8), 749–757 (2000).
Eichenlaub, M. and Redish, E. F., *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer, 2019), pp. 127–151.
Elby, A., Am. J. Phys. **69**(S1), S54–S64 (2001).
Etkina, E., Phys. Rev. Spec. Top. Phys. Educ. Res. **6**(2), 1–26 (2010).
Etkina, E. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(1), 10127 (2018).
Etkina, E. *et al.*, Phys. Teach. **44**(1), 34–39 (2006).
Evans, V., *Cognitive Linguistics* (Edinburgh University Press, 2006).
Fauconnier, G., *The Analogical Mind: Perspectives From Cognitive Science* (The MIT Press, 2001).
Fauconnier, G. and Turner, M., *The way we Think: Conceptual Blending and the Mind's Hidden Complexities* (Basic Books, 2008).
Feynman, R., *The Character of Physics Law* (MIT Press, 1965).
Forjan, M. and Slisko, J., Eur. J. Phys. Educ. **5**(3), 20–32 (2014).
Freitas, I. M. *et al.*, Res. Sci. Educ. **34**(1), 113–133 (2004).
Gentner, D., Cognit. Sci. **7**(2), 155–170 (1983).
Geyer, M.-A. and Kuske-Janßen, W., *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer, 2019), pp. 75–102.
Giere, R. N., *Understand Scientific Reasoning* (Renehart and Winston, Inc, FL Holt, 1991).
Giere, R. N., *Model-based Reasoning in Scientific Discovery* (Springer, 1999), pp. 41–57.
Giere, R. N., Argumentation **15**(1), 21–33 (2001).
Giere, R. N., Philos. Sci. **71**(5), 742–752 (2004).
Gingras, Y., History Sci. **39**(4), 383–416 (2001).
Gingras, Y., Sci. Educ. 24, 529–541 (2015).
Gire, E. and Price, E., Phys. Rev. Spec. Top.: Phys. Educ. Res. **10**(2), 1–11 (2014).
Greca, I. M. and de Ataíde, A. R. P., Springer Proc. Phys. **190**, 55–64 (2017).
Gupta, A. and Elby, A., Int. J. Sci. Educ. **33**(18), 2463–2488 (2011).
Halloun, I. I. A., *Modeling Theory in Science Education* (Springer, 2006).
Halloun, I. I. A., Sci. Educ. **16**, 7 (2007).
Hammer, D., Cognit. Instr. **12**(2), 151–183 (1994).
Hammer, D., Am. J. Phys. **68**(S1), S52–S59 (2000).
Hammer, D. and Elby, A., *Personal Epistemology: The Psychology of Beliefs About Knowledge and Knowing* (Routledge, 2002), pp. 169–190.
Hammer, D. and Elby, A., J. Learn. Sci. **12**(1), 53–90 (2003).
Hansson, L. *et al.*, Sci. Educ. **24**, 615–644 (2015).
Heck, A. and Buuren, O. V., *Mathematics in Physics Education* (Springer, 2019), pp. 53–74.
Hestenes, D., Am. J. Phys. **55**(5), 440 (1987).
Hestenes, D., AIP Conf. Proc. **399**, 935–958 (1997).
Hewitt, P. G., *Conceptual Physics* (Pearson, 2002).
Hewitt, P. G., Phys. Teach. **49**(5), 264 (2011).
Hrepic, Z. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **6**(2), 1–18 (2010).
Hu, D. and Rebello, N. S., Phys. Rev. Spec. Top.: Phys. Educ. Res. **9**(2), 1–15 (2013).
Janssen, W. and Pospiech, G., Verbalization of formulas-The meaning of formulas and their mediation. PhyDid B didactics of physics contributions to the DPG spring conference (2015).
Jensen, J. H. *et al.*, Int. J. Math. Educ. Sci. Technol. **48**(1), 1–15 (2017).
Jewett, J. W., Phys. Teach. **46**(1), 38–43 (2008).
Karam, R., Phys. Rev. Spec. Top.: Phys. Educ. Res. **10**(1), 1–23 (2014).
Karam, R. and Krey, O., Sci. Educ. **24**, 661–698 (2015).
Kim, M. *et al.*, New Phys.: Sae Mulli **66**(1), 50–60 (2016a).
Kim, M. *et al.*, New Phys.: Sae Mulli **66**(1), 61–71 (2016b).
Kim, M. *et al.*, J. Korean Phys. Soc. **73**(2), 145–151 (2018).
Kim, M. *et al.*, New Physics: Sae Mulli **70**(10), 851–862 (2020).
Kirschner, S. *et al.*, Int. J. Sci. Educ. **38**(8), 1343–1372 (2016).
Kjeldsen, T. H. and Lützen, J., Sci. Educ. **24**(5–6), 543–559 (2015).
Kneubil, F. B. and Robilotta, M. R., Sci. Educ. **24**(5–6), 645–660 (2015).
Knight, R. D., *Five Easy Lesson Strategies for Successful Physics Teaching* (Addison Wesley, 2004).
Kragh, H. (2015). arXiv:1510.04046, pp. 1–20.
Krauss, S. *et al.*, ZDM Int. J. Math. Educ. **40**(5), 873–892 (2008).
Kuo, E. *et al.*, Sci. Educ. **97**(1), 32–57 (2013).
Lakoff, G. and Johnson, M., Cognit. Sci. **4**(2), 195–208 (1980).
Lakoff, G. and Nunez, R., *Where Mathematics Comes From* (Basic Books, 2000).

Lavagnini, M. *et al.*, J. Phys.: Conf. Ser. **1929**(1), 012082 (2021).
Lederman, N. G. *et al.*, J. Res. Sci. Teach. **39**(6), 497–521 (2002).
Lehavi, Y. *et al.*, Springer Proc. Phys. **190**, 95–104 (2017).
Lehrer, R. and Schauble, L., *Cultivating Model-Based Reasoning in Science Education* (Cambridge University Press, 2006).
Lloyd, G. and Sivin, N., *The way and the Word: Science and Medicine in Early China and Greece* (Yale University Press, 2002).
Lochdead, J., Educ. Leadership **39**(1), 68–70 (1981).
Lopez-Gay, R. *et al.*, Sci. Educ. **24**, 591–613 (2015).
Mach, E., *The Science of Mechanics: A Critical and Historical Account of Its Development* (Open Court Publishing Company, La Salle, 1989).
Magnusson, S. J. *et al.*, *Examining Pedagogical Content Knowledge* (Springer, 1999), pp. 95–132.
Malone, K. L., Phys. Rev. Spec. Top.: Phys. Educ. Res. **4**(2), 1–15 (2008).
Marrongelle, K. A., School Sci. Math. **104**(6), 258–272 (2004).
Mason, A. J. and Singh, C., Eur. J. Phys. **37**(5), 055704 (2016).
McComas, W. F., *The Nature of Science in Science Education* (Springer, 2002), pp. 53–70.
Meli, K. *et al.*, Can. J. Sci., Math. Technol. Educ. **16**(1), 48–63 (2016).
Michelsen, C., ZDM Int. J. Math. Educ. **38**(3), 269–280 (2006).
Michelsen, C., Phys. Educ. **50**(4), 489–494 (2015).
Modir, B. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 1–12 (2017).
Mortimer, E. F., Sci. Educ. **4**(3), 267–285 (1995).
Mulhall, P. and Gunstone, R., Res. Sci. Educ. **38**(4), 435–462 (2008).
Niaz, M., J. Sci. Educ. Technol. **8**(2), 145–150 (1999).
Niess, M. L., Teach. Teach. Educ. **21**(5), 509–523 (2005).
Niss, M., Int. J. Sci. Math. Educ. **15**(8), 1441–1462 (2017).
Park, S. and Oliver, J. S., Res. Sci. Educ. **38**(3), 261–284 (2008).
Pietrocola, M., Mathematics as structural language of physical thought. Connecting research in physics education with teacher education, 2 (2008).
Planinic, M. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **9**(2), 1–9 (2013).
Pospiech, G., *Mathematics in Physics Education* (Springer, 2019), pp. 1–36.
Pospiech, G. and Fischer, H. E., *Physics Education* (Springer, 2021), pp. 201–229.
Pospiech, G. and Oese, E., *Active Learning—In a Changing World of new Technologies* (Charles University in Prague, MATFYZPRESS Publisher, 2014), pp. 199–206.
Pospiech *et al.*, *Mathematics in Physics Education* (Springer, 2019).
Rebello, S. and Cui, L., *2008 Annual Conference & Exposition* (Chicago, 2008), pp. 13–1048.
Redish, E. F., arXiv preprint physics/0608268 (2006).
Redish, E. F., Springer Proc. Phys. **190**(September), 25–40 (2017).
Redish, E. F., Phys. Teach. **59**(5), 314–318 (2021).
Redish, E. F. and Kuo, E., Sci. Educ. **24**(5–6), 561–590 (2015).
Reiner, M. *et al.*, Cognit. Instr. **18**(1), 1–34 (2000).
Roorda, G. *et al.*, Int. J. Sci. Math. Educ. **13**(4), 863–889 (2015).
Sands, D., *Concepts, Strategies and Models to Enhance Physics Teaching and Learning* (Springer, 2019), pp. 73–86.
Scherr, R. E. and Hammer, D., Cognit. Instr. **27**(2), 147–174 (2009).
Serway, R. A. and Vuille, C., *College Physics* (Brooks/Cole, Boston, 2012).
Sherin, B. L., Cognit. Instr. **19**(4), 479–541 (2001).
Shulman, L. S. (1986). Definición de cómputo - Qué es, Significado y Concepto. American Educational Research Association Is Collaborating with JSTOR to Digitize, Preserve and Extend Access to Educational Researcher, **15**(2), 1.
Siswono, T. Y. E. *et al.*, J. Phys.: Conf. Ser. **812**(1), 012046 (2017).
Song, J. *et al.*, Phys. Teach. Educ. Beyond, 359–366 (2000).
Spurgin, C. B., Phys. Educ. **19**, 114 (1984).
Toulmin, S., *The Philosophy of Science* (Hutchinson, 1953).
Tuminaro, J. and Redish, E. F., Phys. Rev. Spec. Top.: Phys. Educ. Res. **3**(2), 1–22 (2007).
Uhden, O., Z. Didaktik Naturwissenschaften **22**(1), 13–24 (2016).
Uhden, O. *et al.*, Sci. Educ. **21**(4), 485–506 (2012).
Uhden, O. and Pospiech, G., GIREP-EPEC Conf. Front. Phys. Educ., 26–31 (2009).
Vogel, M. *et al.*, Comput. Educ. **49**(4), 1287–1298 (2007).
Walker, J., *Fundamental of Physics* (John Wiley & Sons, Inc., 2014).
Zahar, E., Br. J. Philos. Sci. **31**(1), 1–43 (1980).
Zalta, E. N., *The Stanford Encyclopedia of Philosophy* (The Metaphysics Research Lab Philosophy Department Stanford University Stanford, n.d.), CA 94305-4115.
Zandieh, M. *et al.*, J. Math. Behav. **33**, 209–229 (2014).

CHAPTER 20 GRAPHS

Lana Ivanjek, Maja Planinic, and Ana Susac

Ivanjek, L., Planinic, M., and Susac, A., "Graphs," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 20-1–20-24.

20.1 INTRODUCTION

Competence to read graphs is not only important in physics but also in other sciences. It is essential for understanding today's world and is one of the main elements of scientific literacy (Glazer, 2011). Researchers and instructors recognize the importance for students to develop an understanding of graphs so that they can interpret scientific facts, analyze data and identify patterns (Berg and Boote, 2017). Students and the general public are exposed to graphs in medical and financial reports, weather and climate patterns, polls, advertisements and daily news. The activities of data handling and those related to graphing according to Gal (2002) may happen in two main contexts: inquiry and reading, where in inquiry context researchers engage in the empirical investigation of actual data and results, make inferences from them and report findings and conclusions, while the reading context happens in everyday situations in which individuals see and interpret graphs in the newspapers, on the TV, in advertisements and on the Internet. Graph skills are thus important for general scientific literacy.

This chapter summarizes the research on student understanding of graphs. It starts with the theoretical background and proceeds to graphs in physics and mathematics education research, describes instruments for assessing student difficulties with graphs and gives the synthesis of the research on student understanding of graphs in mathematics, physics and in real life problems. At the end, the eye-tracking studies related to student understanding of graphs are presented and the implications of all research findings for teaching and learning of graphs are discussed. Because of the large number of research papers on this topic is the overview of literature not exhaustive.

20.2 THEORETICAL BACKGROUND

Graphs are very commonly used in science and mathematics as well as in the teaching of those subjects. It is often assumed that they make quantitative information easier to understand. Research

https://doi.org/10.1063/9780735425514_020

The International Handbook of Physics Education Research: Special Topics, Mehmet Fatih Taşar and Paula R. L. Heron

shows, however, that students have many difficulties with graph interpretation in both physics and mathematics (e.g., McDermott *et al.*, 1987; Leinhardt *et al.*, 1990; Beichner, 1994; Planinic *et al.*, 2013; and Ivanjek *et al.*, 2016). In their review article, Shah and Hoeffner (2002) identify three components of graph comprehension from the literature. First, viewers must encode the visual array and identify the important visual features of the graph. Second, viewers must relate the visual features to the conceptual relations that are represented by those features. Third, viewers must determine the referent (e.g., speed of car A or car B in a race) of the concepts being quantified and associate those referents to the encoded functions. It seems therefore that three factors play an important role in graph comprehension: the characteristics of the visual display, knowledge about graphs and content of the graph (Shah and Hoeffner, 2002).

1. *Visual display*. The type of graph that is used should match its purpose—line graphs are good for depicting x-y trends, bar graphs for discrete comparisons and pie charts for relative proportions. Color in graphs can provide helpful cues and reduce viewers' difficulty in keeping track of graphic referents (Carpenter and Shah, 1998) or enable them to group data easier (Lewandowsky and Spence, 1989). Graph designers should avoid using legends and instead label graph features directly with their referents (Kosslyn, 1994) and generally avoid any noninformative features that may be distracting.
2. *Knowledge about graphs*. Viewers' knowledge and expectations may affect how they encode and remember graphs (Shah and Hoeffner, 2002). In the case of line graphs, viewers tend to favor 45° lines (Schiano and Tversky, 1992) and expect a steeper line to always represent faster rates of change, irrespective of how variables are plotted (Gattis and Holyoak, 1996). If they are prone to slope-height confusion, they will attempt to infer the slope from the local height of the graph (Leinhardt *et al.*, 1990).
3. *Knowledge about content*. Viewers' knowledge and expectations about the content of the graph may also affect their interpretation. Viewers tend to infer a relationship from the graph that they expect or to mistake a graph for the literal picture of the situation.

A graph is a symbolic representation of the relationship between variables and its processing requires the ability to perceive and remember a pattern of spatially arranged visual data and the ability to reason for spatial visual information. Some studies have suggested that understanding graphs is related to logical thinking (Berg and Phillips, 1994), spatial ability and mathematics achievement (Bektasli and White, 2012), and that constructing and interpreting graphs requires formal operational reasoning (Wavering, 1989; and Beichner, 1990). Other researchers emphasized that graphing ability is influenced mostly by practice (or lack thereof) instead of reflecting a cognitive ability (Roth and McGinn, 1997). One line of research investigated the role of visual-spatial ability in the graph interpretation. Spatial ability can be defined as the ability to generate, retain, retrieve, and transform well-structured visual (mental) images (Lohman, 1996). Kozhevnikov *et al.* (2007) found a significant relationship between students' visual-spatial ability and their solutions to kinematics problems, some of which included graphs. High- and low-spatial students were more likely to interpret kinematic graphs as abstract representations of an object's motion, whereas high- and-spatial students were more likely to interpret graphs as picture-like representations. However, it seems that the correlation between spatial ability and kinematics problem solving is no longer present after students receive physics instruction (Kozhevnikov and Thornton, 2006). This suggests that when conceptual knowledge has already been developed, spatial ability is no longer a predictor of their performance on the kinematics problems. However, spatial ability and conceptual knowledge may not be alternative explanations of the differences in students' performance but may be interrelated in a way that high spatial ability may enable students to form better conceptual knowledge in physics (Kozhevnikov *et al.*, 2007). The physics instruction rich in visualization technologies may help students with low spatial ability who have trouble generating visualizations on their own to develop a better understanding of graphs (Kozhevnikov *et al.*, 2007).

It seems that graph comprehension processes are influenced by some characteristics of the students, such as their visual-spatial ability, content knowledge, and formal reasoning level, but also by the characteristics of the graph (its type, content, and visual display). How the process of comprehension unfolds is described by several cognitive models (Pinker, 1990; Carpenter and Shah, 1998; and Shah *et al.*, 2012).

The model of Shah *et al.* (2012) builds on Pinker's (1990) model. It includes both conceptual and perceptual processes, and suggests that the interplay of visual features of the graph, viewer's prior knowledge about the content and their graph reading expertise determine the viewer's interpretation of the data (Shah *et al.*, 2012). Visual characteristics of the graph determine the visual chunks that are encoded by the viewer, which are then processed sequentially (Carpenter and Shah, 1998). When visual chunks are activated, prior knowledge and expectations of the viewer are also activated, which influence their interpretation of the graph (Shah *et al.*, 2012). It is important that the viewer knows what the features of the graph imply about the quantitative relationship depicted by the graph. This is referred to as graph schema (Pinker, 1990). If the viewer possesses the right graph schema, the process of graph interpretation is automatic (e.g., recognizing the linear relationship of the variables from a line graph). However, if a viewer does not possess adequate graph schemas (does not possess the knowledge of certain common graphical patterns and is not able to recognize them), the process of graph comprehension becomes much more complicated and demanding for the viewer, often resulting in mistakes. On the other hand, graph schemas can also distort the remembered graphs (e.g., remembering a line much closer to 45° than it was in reality).

Viewers' content knowledge and the type of content in the graph may also influence their comprehension of the graph. In kinematics graphs, for example, students have problems interpreting rates of change through evaluating slopes of the graphs (Leinhardt *et al.*, 1990; Beichner, 1994; and Planinic *et al.*, 2012) or interpreting areas under a graph (Nguyen and Rebello 2011; Christensen and Thompson, 2012; and Planinic *et al.*, 2013) and they seem to prefer graphs which show time dependence of a variable (Leinhardt *et al.*, 1990).

Even though graphs and diagrams are commonly used in education with the idea that they promote understanding better than other representations, some researchers argue that this cannot be stated generally, one reason being that they are not a well-defined unitary class of representations (Cheng *et al.*, 2001). They state that diagrams and propositions may not be so different since diagrams may include propositions in some cases, and propositions may include some diagrammatic properties to encode information (e.g., in some formulas it matters on which side of the equation a certain symbol is placed). They summarize certain findings of cognitive science on the comprehension of diagrams (Cheng *et al.*, 2001):

- diagrams can be regarded as an arrangement of various graphic elements in space, which can be grouped or distinguished according to the visuospatial characteristics of the display;
- locational indexing in diagrams makes them more effective than informationally equivalent sentential representations, probably because of less search time for the required information;
- the major factor determining a viewer's capacity to make effective use of the diagram is the person's prior knowledge about the content of the diagram and the method of depiction;
- cognitive approaches that we use for interpreting our everyday visual environment are inappropriate for analyzing diagrams;
- skills for interpreting diagrams must be learned and are domain-specific, although there are some generic aspects too; and
- background knowledge of the viewer plays a critical role (key factor is what the user brings to the diagram, not what the diagram brings to the user).

In summary, cognitive science points to some important aspects of graphs and diagrams that can be used in teaching and learning physics as well as science in general. Graphs and diagrams can be of great help for the representation of scientific data, but caution is required in their use with novices who may not have the necessary skills or background content knowledge to extract information from them efficiently and are at the same time very much influenced by the type of graph/diagram used and its visual characteristics. Their degree of formal reasoning development and their visual–spatial ability may also contribute to their success or problems in constructing and interpreting graphs and diagrams. As Carpenter and Shah (1998) state, it is important to bear in mind that "relatively simple graphical displays may require relatively complex cognitive processes." It is therefore essential to help students develop the necessary skills as well as the necessary content knowledge and to integrate and use those skills and knowledge often for graph interpretation and graph construction in physics teaching.

20.3 GRAPHS IN MATHEMATICS EDUCATION RESEARCH

20.3.1 Interpretation of graphs

Most of the graphs that students encounter in mathematics are not just the purely mathematical functional dependencies of y-variable on x-variable, but often involve additional contexts, such as, for example, representations of the changes of, e.g., temperature, population, position, distance, or speed over time.

When interpreting the graph, two groups of activities take place: identification and reading. The identification process includes extracting information without looking at the actual data points. The focus is on the diagram frame and coordinate axes. The following activities take place during the identification process: assignment of the quantities to the axes, determination of the scale range and, if several data series are shown, the assignment of data series. The reading process consists of reading the values from the graph and extracting functional dependencies and correlations. Each of these activities may be a source of possible difficulties with graph interpretation, as will be shown in this chapter.

The first extensive study on students' understanding of graphs was done by Kerslake (1977, 1981). The study with about 1400 British school students aged 13–15 has shown that most of the pupils (90%) could successfully read off or plot the points in the graph, but that using decimals in data points reduced the success rate to 70%. The problems emerged when students were asked about the graph slopes (30%–35% of success rate) and only about 5%–10% of students recognized that parallel lines correspond to equal slopes. The relation between straight lines and their equations was understood by 5%–30% of students, depending on their age. This research was followed by the Bell and Janvier's research on interpreting graphs in 1981 (Bell and Janvier, 1981). Students were asked during interviews to interpret graphs in different contexts starting from the increase in weight for boys and girls over time, to populations of microbes in relation to the time of feeding and height of the plants over time, to the speed of racing cars over time. The presented graphs did not present a linear dependence but were curved graphs. When interpreting graphs, the following student difficulties emerged:

- not distinguishing between the greatest value and the greatest increase;
- confusion among the rate of decrease (gradient), amount of increase (interval) and the greatest value;
- confusion between the interval and the point;
- difficulty interpolating values and answering about the value which is between given points;
- pictorial distraction, where the shape of the graph was confused with the race track being traveled and difficulties in the interaction between graph; and
- situation where students had difficulty in extracting the meaning from the graph.

Curcio (1987) noted that the understanding of graphs is affected by different factors such as students' prior knowledge, the mathematical content embedded in the graph and the form of the graph. From there, he defined three levels of graphical interpretation ability (1) reading the data, or the ability to obtain information directly from the graph; (2) reading between the data, or the ability to identify relationships in the graph data; and (3) reading beyond the data, or the ability to make inferences and predictions from the graph data, where the third level has been shown to be most difficult for students. Students have fewer difficulties when "reading the data," but they make errors in "reading between the data" questions (Dossey *et al.*, 1993; and Zawojewski and Heckman, 1997).

In 1990, Leinhardt, Zaslavsky, and Stein conducted a literature review on the construction and interpretation of graphs and functions (Leinhardt *et al.*, 1990). They have summarized the difficulties found both in mathematics education research and in science education research when students were confronted with graphs. Difficulties with interpretation included interval/point confusion, slope/height confusion, and iconic interpretations. The slope-height confusion is evident when students mistakenly replace the slope with the local height value of the graph. This difficulty might be an obstacle when analyzing pattern changes that are time dependent. Interval-point confusion is evident when students focus on a single point rather than on an interval (Leinhardt *et al.*, 1990). Considering the graph as a picture is present when students are unable to treat the graph as an abstract representation of a relationship but consider it as a picture of a particular situation. For example, if there is an increasing straight line in a distance vs time graph, a student who sees the graph as a picture will interpret the graph as representing a body moving up an incline.

In their study with undergraduate students, Shah and Freedman (2011) found associations between readers' interpretation of graphs and the format of the graph (line or bar graph) and between the readers' familiarity with the content of the graph, and their understanding of graphs.

Recent studies also examined the pre- and in-service teachers. Studies of Arteaga *et al.* (2015), Jacobbe and Horton (2010), and Patahuddin and Lowrie (2019) have utilized Curcio's theory on graphical interpretation (Curcio, 1987). They found that the teachers also had problems with interpreting graphs at the highest level of "reading beyond data." Patahuddin and Lowrie (2019) found that 90% of the Indonesian teachers were successful in the task involving reading the data, while more than 70% of the teachers had a problem with reading beyond the data and most of the teachers interpreted the graph as an iconic representation of a real event or treated the graph as a picture.

20.3.2 Construction of graphs

In the construction of graphs, a distinction can be made between two main activities: the construction of the frame and the data entry (Geyer, 2019). The construction of the frame refers to the coordinate axes, scaling of the axes, and labels. When the students are given the data from an experiment or some other source, they usually need to perform the following steps: selection of the relevant variables, choice of quadrants in a Cartesian coordinate frame, assignment of the variables to the axes, drawing the scales on the axes, adding a legend or a label and selecting a diagram type (for example a line diagram or a histogram). Data entry includes entering the points into a diagram (pairs of measures) and displaying dependencies and correlations (for example with a best fit line or some other function). Each of these steps is a possible source of student difficulties that were investigated in mathematics education research starting from 1977 (Kerslake, 1977), but looking at the graphs as a separate topic was first mentioned in 1966 by Knight (1966), where it was noted that there are pupils with different abilities for drawing graphs.

Difficulties in graph construction may arise already during the construction of the coordinate axes when students need to select the relevant variables. They sometimes interpret the given values in a time frame and introduce a time axis, even if no values are given for it (Brasell and Rowe, 1993). When assigning the variables to the axes, students may reverse independent and dependent variables (Brasell and Rowe, 1993). For some students, the difference between the dependent and the independent variables is unclear (Pospiech *et al.*, 2011). As the next problem, the labeling of the axes may be partly non-existent or incomplete; for example, the physical units may be forgotten (Brasell and Rowe, 1993). Similar results were also found in the biology education study by Lachmayer (2008), in which in 10%–23% of the cases the assignment of the quantities to the axes did not correspond to the convention, and the dependent and independent variables were interchanged. Lachmayer also observed the missing or incomplete labeling of the axes.

Difficulties in the choice of quadrants were observed in a mathematics education study. Here, students drew only the first quadrant, even though negative x- and y-values should also be entered. In part, these negative values were integrated on the positive axis (Kerslake, 1981). Furthermore, difficulties occurred with the scaling of the coordinate axes. In the study of Brasell and Rowe (1993), some students adapted the axes and scales to the available paper format and did not make any content considerations for them. In a purely mathematical diagram construction, some students had the idea that the positive and negative parts of an axis could be scaled differently or that the scale does not start at zero but at the smallest positive value of the given pairs of values (Kerslake, 1981). Furthermore, in the case of purely mathematical diagrams, students sometimes assumed that the scaling of the x- and y-axes had to be symmetrical, even if this made the functional properties no longer recognizable (Goldenberg, 1988).

Students may also have difficulties entering and connecting data points. Some students enter values incorrectly into a graph or forget individual data points (Brasell and Rowe, 1993). When entering pairs of values into a graph without context, 13- to 15-year-old students had more difficulty with non-integer values than with integer values. After the points have been entered in the graph, it is not always clear for students whether they can draw a function graph or not. They connect point to point with straight lines. The connecting lines are seen only as the connecting elements, but no conceptual meaning is attributed to them (Leinhardt *et al.*, 1990; Brasell and Rowe, 1993; and Glazer, 2011). Even undergraduate university students sometimes have difficulties in drawing a regression curve and seem to apply only a not well-understood algorithmic procedure. The relation to the physics involved is hardly established (Nixon *et al.*, 2016). When deciding whether a continuous function graph can be drawn, students focus partly on the appearance of the discrete points. A reference to the context is not recognizable in these cases (Kerslake, 1981). It could be shown that in an abstract or conceptualized diagram construction students often tend to draw a linear graph. This may be mainly due to the fact that linear relationships are treated in class first, before other types of functions are introduced (Leinhardt *et al.*, 1990; and Mevarech and Kramarsky, 1997). The problem of linearity or the tendency to always construct a straight line was also presented in other studies (Markovits *et al.*, 1983; and Dreyfus and Eisenberg, 1987).

The graph-as-picture difficulty can also be observed in the construction of diagrams and not only in their interpretation. Students sometimes do not represent a functional relationship between two quantities but have a pictorial, iconic idea of a function graph. For example, they represent a given drop and bounce heights of a ball as a process pictorially, as one might observe it in a video recording (Brasell and Rowe, 1993). Likewise, the slope-height-confusion could also be observed when constructing diagrams. Students confuse the slope of a graph the minimum or maximum values (Leinhardt *et al.*, 1990).

In the study of Mevarech and Kramarsky (1997), eighth graders were asked to represent the everyday situations, which were described verbally, with graphs. When sketching those graphs they constructed, for example, only one point for a described situation or represented the change of the two described variables in one diagram each (Mevarech and Kramarsky, 1997). The tendency towards linear function graphs was observed as well, where increasing functions were preferred over decreasing ones. In some cases, the students reversed the axis scaling so that the graph still looked increasing (Mevarech and Kramarsky, 1997). In a similar study by Hattikudur *et al.* on linear functions in everyday contexts, students had particular difficulty in correctly mapping the y-intercept (Hattikudur *et al.*, 2012). The slope posed fewer problems. Graph construction was more difficult for the students when the graph was presented with qualitative features than with quantitative features (Hattikudur *et al.*, 2012). In addition, the constant function represents a particular problem. Some students drew it as a linearly increasing graph but adapted the scaling in such a way that the same value was always plotted on the axis. Other students drew several increasing straight lines (Mevarech and Kramarsky, 1997).

20.4 GRAPHS IN PHYSICS EDUCATION RESEARCH

One of the first studies on student understanding of graphs in physics is the one from 1987 by McDermott, Rosenquist, and van Zee on student difficulties in connecting graphs and physics (McDermott *et al.*, 1987). In their work, two different tasks regarding graphs in physics were investigated: interpretation and construction of graphs. When asked to interpret a graph, students need to make sense of the graph and gain meaning about the physics content or some other subject content from it. On the other hand, construction consists of creating something new from the description or the given data, for example, creating a graph or plotting the points from the measurement data. Although interpretation and construction of graphs are not mutually exclusive, the research on graphs will be presented based on these two aspects. Both interpreting and constructing (drawing) graphs are of critical importance for the development of understanding of different physics topics, especially kinematics. This is probably the reason why most of the research on student understanding of graphs includes topics from kinematics (e.g., McDermott *et al.*, 1987; Mokros and Tinker, 1987; Brasell and Rowe, 1993; Beichner, 1994; and Wemyss and van Kampen, 2013).

20.4.1 Interpretation of graphs

In 1987, McDermott *et al.* collected data on tasks on connecting graphs to physical situations and to the real world from students at the University of Washington, enrolled in preparatory physics courses for undergraduate students, a special course for preservice teachers and standard introductory courses

(McDermott *et al.*, 1987). Although there was some prior research on student understanding of graphs in mathematics (Kerslake, 1977; and Bell and Janvier, 1981), this was the start of the systematic research on student understanding of graphs in physics. One of the main difficulties found was the difficulty in discriminating between the slope and the height of the graph. Students were given a position-time graph for the motion of two objects, and they were asked if at some instant the objects had the same speed. To answer that question, students only needed to recognize that the slopes of the lines represent the speeds of the objects and then to compare the slopes. But instead of that, many students concentrated on the height of the graphs at one instant and compared them. Further, students had difficulty interpreting the changes in height and changes in slope, which was evident when the question involving curved graphs was posed. Curved graphs are more complex than line graphs because besides comparing and calculating slope, students need to focus on changes in heights and changes in slope which are not necessarily the same: for example, the height of the graph could increase and at the same time the slope could decrease. Other difficulties included looking at the sign of the slope instead of the changes in the magnitude of velocity when deciding whether the object is speeding up or slowing down, problems relating one type of graph to another where students had difficulty realizing that the slope of the position-time graph corresponds to the height of the velocity-time graph, inability to visualize the motion that is represented in a velocity-time graph and difficulties interpreting the area under the graph.

Approximately at the same time, in a study on the impact of microcomputer-based labs on children's ability to interpret graphs, Mokros and Tinker (1987) also found two types of student errors: slope-height confusion and treating the graph as a picture of motion. They suggested that students find the incorrect picture-like graph so visually compelling that they select it without really thinking about the other options. However, students showed significant gains on graphing items after a microcomputer-based laboratory that, according to the authors, uses multiple modalities, pair events with their graphical representations in real time, provides genuine scientific experiences and eliminates the drudgery of graph production.

Most of the difficulties found by McDermott *et al.* were confirmed by subsequent research. In 1994 Beichner (1994) constructed the test on student interpretation of kinematic graphs and the main difficulties identified in the analysis were graph as a picture (where students see the graph as the image of the motion), the confusion of variables (not distinguishing between position, velocity and acceleration and thinking that these graphs need to be the same), difficulty determining slope (the most common error being the dividing of the y-value with x-value to determine speed of points on a line graph not going through the origin), slope-height confusion (extracting the information from the height of the graph when slope is required) and confusion between slope and area under the graph (a tendency to calculate slope instead of area). The student difficulty of interpreting a graph as a picture of motion was already found by Kerslake (1977) in mathematics education research, when students were presented with a distance vs time graph consisting of increasing and decreasing lines and said that the graph shows "climbing a mountain: first going uphill, then going downhill" or "climbing a vertical wall" and "going east, then north and then east."

One year earlier, Brasell and Rowe (1993) gave questions about graphs to students in 12th grade in North Florida. In seven items that focused on constant velocity, students were either given the verbal description of a motion and needed to select the appropriate section of the graph or were given a graph and asked to select the appropriate verbal description. The authors concluded that changes in verbal descriptions (use of colloquial language for example "moving steadily" vs use of scientific language, such as "constant velocity") can have a big impact on the percentage of correct answers. Also, when the variable is described, students only need to translate the verbal description to a graphic representation, whereas when the event is described, students have one more step to do because they first need to extract the relevant information about the variable, which makes the task more complex. Brasell and Rowe (1993) also concluded that student difficulties depend on the direction of translation between two representations. In their research, the error rates were much higher (by about 30%) for items starting with a verbal description of an event that needed to be matched with one of the four lines in the graph (verbal to graphical translation) than for the items starting with the graph that needed to be matched with one of the four verbal descriptions (graphical to verbal translation). They have compared this with the translation between languages, where it is easier to translate from a less familiar language to a more familiar language, which corresponds to a translation from graphs to verbal description.

In some newer studies, Wemyss and van Kampen (2013), Planinic *et al.* (2013), Bollen *et al.* (2016), and Ivanjek *et al.* (2016) have also investigated student understanding of graphs. Wemyss and van Kampen (2013) found that students have difficulties with linear distance-time graphs, including the difficulty of determining the direction of the motion from a graph. Only 20% of students correctly determined the value of the speed from a linear distance-time graph, with 50% of students just dividing the coordinates in order to determine the speed. Planinic *et al.* (2013) and Ivanjek *et al.* (2016) found that university students' understanding of kinematic concepts is still not sufficiently developed, that their preferred strategy to solve physics questions is the use of the formulas (often incorrect ones), that calculating the slope is the most difficult aspect in the slope questions and that interpretation of the meaning of the area under the graph is very difficult for the students.

All studies mentioned up to now focus on kinematics. In 2004, Forster looked at questions with a graphical component on Tertiary Entrance Examinations in Western Australia (Foster, 2004). The questions related to topics other than mechanics, such as "Sound wave", "Electric power," "Structures and material," and one on "movement. In summary, the author concluded that non-success in graphing questions had a source in students" non-familiarity with phenomena, physics principles and definitions. Other obstacles were the difficulty to draw lines of best fit, not reading scales accurately, not paying attention to scales in construction, and difficulties with gradient-, slope-height-, and interval-point confusion.

20.4.2 Construction of graphs

There are not so many studies on the graph construction as on their interpretation (Glazer, 2011; and Nixon *et al.*, 2016). Most existing studies do not cover solely the interpretation or the construction of graphs but contain separate questions on both skills. McDermott *et al.* (1987) asked students to draw the graphs for the motion of a steel ball rolling on different combinations of ramps. The main difficulties they found were (1) failing to distinguish between the position of the ball at the particular moment and the displacement of the ball during time interval (students drew a point instead of a straight line in the position-time graph for a stationary object), (2) connecting the shape of the graph with the path of motion (graph as a picture), (3) difficulty representing negative velocities on velocity-time graph, (4) not being aware that one cannot tell only from the acceleration-time graph if the object is speeding up or slowing down and in which direction it is traveling, and (5) difficulty distinguishing between different types of motion graphs and accepting the idea that the same motion can be represented with graphs of different shapes.

Brasell and Rowe (1993) found that 12th grade student (aged 17–18) students have a range of difficulties when asked to construct a graph representing the bouncing ball data. Difficulties ranged from missing the understanding of how Cartesian graphs represent data to reversing axes (transposing the dependent and independent variables) using unbalanced scales and connecting the points instead of drawing the best-fit line. The treatment of graphs as pointwise and connecting the points rather than applying the best-fit line was already reported by Padilla *et al.* in 1986.

Nixon *et al.* investigated university students' construction of graphs in the context of two physics lab activities (Nixon *et al.*, 2016). Undergraduate students in their study could successfully construct graphs with best fit lines but had problems relating graphs to the underlying physics concepts. The most common strategy students used to draw the best-fit line was to split the data in half, so that half of the points were above the line and half of the points were below the line. This strategy was followed by the strategy to draw a line down the middle of the data points and with getting the line as closely as possible to the maximum number of data points. Overall, students provided high-quality graphs and best fit lines. This finding contradicts the previous findings from Brasell and Rowe (1993), and the main reason could be the higher age of students. When asked to interpret the best fit line, students showed understanding of best-fit lines primarily in terms of their procedural value, to mitigate error, show the connection between variables, and calculate a value—rather than in terms of the physics concepts they represent.

Geyer described in her dissertation an exploratory laboratory study conducted with 17 pairs of students aged about 14 years, in which they were asked to work on tasks in the field of thermodynamics. These tasks contained (besides other representations) the construction of a graph starting from a table, a formula, or a verbal description (Geyer, 2019; and Pospiech *et al.*, 2019). The main difficulties students had with graph construction were as follows:

- difficulties with the selection of relevant quantities to put on the axes when more than two quantities are involved and the preference for time dependence;
- difficulties in assigning the quantities to the axes: using (incorrectly) remembered rules and difficulties determining (in)dependent variables;

- difficulties with choosing the quadrants;
- difficulties with scaling the axes;
- difficulties representing functional dependencies and relationships, where students were sketching the curve connecting all data points or were incorrectly assuming the type of function; and
- difficulties with extrapolation.

Pospiech and Geyer concluded that the observed students generally showed an algorithmic stepwise approach to graph construction, but depending on the task also used characteristics of the functional relation or tried to verify their solutions, which showed that even some students in their third year of school physics were able to use more advanced strategies (Pospiech *et al.*, 2019).

20.5 INSTRUMENTS FOR ASSESSING STUDENT DIFFICULTIES WITH GRAPHS

The most widely used instrument for measuring graph comprehension is the Test of Understanding Graphs in Kinematics (TUG-K) developed by Beichner (1994). The initial version of the TUG-K was developed based on seven objectives that students are usually expected to achieve during introductory university courses in mechanics:

1. Determine the velocity from the position-time graph.
2. Determine the acceleration from the velocity-time graph.
3. Determine the change of position in an interval from the velocity-time graph.
4. Determine the change of velocity in an interval from the acceleration-time graph.
5. Select another corresponding graph.
6. Select a textual description from a graph.
7. Select a graph from a textual description.

Objectives 1 and 2 relate to the concept of slope, objectives 3 and 4 examine the understanding of the area under the graph, while objectives 5–7 refer to both slope and area under the curve. Three multiple-choice questions were designed for each objective so that the TUG-K contains 21 items.

The initial version of the test was iteratively administered to over 350 high school and college students and subsequently revised. The content validity was established by giving the test to 15 instructors, including high school, college, and university faculty. The final version of the TUG-K was administered to more than 500 college and high school students. The analysis of the item difficulty, the discriminatory power, and the reliability showed that most items had adequate discrimination and the overall reliability of the TUG-K was good.

For many years, the TUG-K was used in physics education research for various purposes. For example, it was used to evaluate the effect of video motion analysis (Beichner, 1996), computational modeling activities (Araujo *et al.*, 2008), a tutorial-type activity (Torres and Alarcon, 2012), and the flipped classroom (Cagande and Jugar, 2018) on student understanding of kinematic graphs. Maries and Singh (2013) investigated the pedagogical content knowledge of the graduate students by asking them which incorrect option would be most often chosen by introductory physics students. Bektasli and White (2012) found that students' skill to determine the slope in a kinematics graph was significantly correlated with logical thinking and gender, but the corresponding correlation to students' skills to determine the area under the graph was not found.

After more than two decades of using the TUG-K, the author and his colleagues decided to make a modified version that would allow comparisons of students' performance on different objectives (Zavala *et al.*, 2017). For example, the differences between the statements of the items of objectives 3 and 4 did not permit direct comparison between the students' ability to determine the change of position from a velocity-time graph (objective 3) and the ability to determine the change of velocity from the acceleration-time graph (objective 4). In the modified version of the TUG-K, some items were changed to allow comparisons of students' performance on different objectives, and the distractors of some of the original items were revised to include the most frequent alternative conceptions. Again, an iterative process of administering the test and its revisions was performed as well as a detailed analysis of the final modified version of the test. The new items showed satisfactory difficulty, discriminatory power, and reliability, and the revised distractors were popular. Overall, the new version of the test had adequate reliability and discriminatory power and it was presented in the PhysPort project (physport.org).

Inspired by the TUG-K, Dominguez *et al.* (2017) developed the Test of Understanding Graphs in Calculus (TUG-C) in a purely mathematical context to evaluate student understanding of the concepts of the derivative as the slope of the tangent to the graph, and the concept of the antiderivative as the area under the graph. The test consists of 16 multiple-choice items that can be found in the appendix of the article in which the TUG-C is presented, and its validity, reliability, and discriminatory power were analyzed (Dominguez *et al.*, 2017). The authors also reported the main students' difficulties with the graphical interpretation of the concepts of the derivative and the antiderivative evaluated in the TUG-C, which could be useful to instructors and researchers in the design of new instructional material.

The relationship between mathematical knowledge of graphs and its application in a physical context is a very active area of research. Planinic *et al.* (2013) developed eight sets of isomorphic questions in mathematics, physics (kinematics), and contexts other than physics. The test was administered to 385 first-year students and the results of the Rasch analysis showed a good functioning of the test. Item and person reliability was satisfactory, as well as the fit of the items with the model evaluated from infit and outfit mean square statistics. A comparison of average difficulties showed that mathematics items were easier for students than both the physics items and the items with contexts other than physics, which suggests that adding the context (physics or other) typically increases item difficulty. Most of the slope items had similar difficulty regardless of the context, whereas items requiring interpretation of the area under the graph were significantly easier in mathematics context than in physics and other contexts. Students were able to calculate the area under the graph in mathematics items, but they struggled to recognize that the same procedure was required in isomorphic physics and other context items. This result suggests that the interpretation of the concept of the area under a graph needs more attention in both physics and mathematics teaching. Further analysis of student explanations of their answers to the test items gave an insight into student reasoning about graphs in different contexts (Ivanjek *et al.*, 2016). The test is available in the supplemental material of the accompanying article (Planinic *et al.*, 2013).

Recently, Carli *et al.* (2020) developed the Test of Calculus and Vectors in Mathematics and Physics (TCV-MP), designed to compare students' ability to answer questions on derivatives, integrals, and vectors in mathematical and physical contexts. The test was administered to more than 1252 students and the obtained reliability and the discriminatory power of the test, both as a whole and at the single-item level, were satisfactory. The test contained 17 pairs of isomorphic questions, and out of the nine pairs related to the understanding of derivatives and integrals, seven contained a graphical representation. On six isomorphic questions regarding graphs in mathematics and physics contexts, students performed better in the mathematics context than in the physics context. The pair of questions in which students had a higher score in physics than mathematics context was not completely isomorphic, so additional analysis was needed for that item. Overall, the results confirmed that knowledge of the necessary mathematical procedures is not enough to solve physics problems.

Another very topical subject in physics education research is the investigation of the use of different representations in physics teaching and learning. The same physical concepts can often be communicated in different forms, e.g., by a graph, equation, diagram, etc. Although the use of multiple representations can strengthen conceptual understanding, it also causes significant difficulties for students. Thus, many PER researchers try to investigate the role of multiple representations, among other things, by developing diagnostic tools for the evaluation of representational competence. Test items containing graphs are often included in instruments that measure representational competence in physics. For example, the Representational Variant of the Force Concept Inventory (R-FCI), Representational Fluency Survey (RFS), and Representational Competence in Kinematics (KiRC) probe student understanding of graphical representation (Nieminen *et al.*, 2010; Hill and Sharma, 2015; and Klein *et al.*, 2017).

Nieminen *et al.* (2010) modified nine original Force Concept Inventory (FCI) items to involve various representations such as graphs, vectors, and motion maps. They found that students' representational consistency (ability to use different representations consistently, correctly or incorrectly, between isomorphic items) considerably depended on the concept, and it increased during the instruction. The Representational Fluency Survey (RFS) was developed to examine representational fluency in physics and administered to university students of different ages (Hill *et al.*, 2014; and Hill and Sharma, 2015). The results showed that the representational fluency improved over the years. The two-tier instrument for representational competence in kinematics (KiRC) developed by Klein *et al.* contained items with formal (mathematical), pictorial and graphical representations, and transitions between them (Klein *et al.*, 2017). It was shown that students with high KiRC scores used representations consistently and changed flexibly between different representations, whereas that was not the case with low-performing students.

One of the important areas of research is the transition from one representation to another. Van den Eynde *et al.* (2019) investigated the transition from a graphical representation to equations and vice versa in mathematics and physics contexts. Responses of students enrolled in an algebra-based and calculus-based physics courses showed that mathematics items are solved better than physics items, and the transition from graph to the equation was easier than from the equation to graph. The authors also analyzed the students' explanations. Students from the calculus-based course used more mathematical arguments and generally scored better on the items.

Furthermore, some PER instruments designed to evaluate students' understanding of certain areas of physics also contain many items with graphs. For example, Lichtenberger *et al.* (2017) developed the kinematics concept test (KCT) to evaluate students' conceptual understanding of kinematics at the high school level, and 32 out of 49 test items contained graphs. Factor analysis of the instrument revealed that students process items with graphs separately from items with pictures and tables, so the authors suggested a more explicit switching between representations in physics teaching and learning.

While most of the above-mentioned instruments were developed and evaluated on senior high school and university students, some instruments were designed so that they could be used on younger high school students as well. The test of graphing in science (TOGS) was developed to assess graphing skills of science students from grades seven through twelve (McKenzie and Padilla, 1986). It was shown that the TOGS was a valid and reliable instrument for measuring skills related to the construction and interpretation of line graphs. Lai *et al.* (2016) developed an instrument to measure middle school students' graph comprehension, critique, and construction in science. Rasch modeling showed that the items formed a coherent scale and had good reliability. Overall, the results indicated that students struggled to link graph features to science concepts, especially when asked to critique or construct graphs.

More recently, Ceuppens *et al.* (2018) developed a 48-item multiple-choice instrument to assess 9th-grade students' representational fluency of linear functions in physics (1D kinematics) and mathematics contexts. The test includes three representations (graphs, tables, and formulas) and six possible representational transitions between them. The results revealed that mean students' scores were significantly lower on physics items than on mathematics items, and students were very successful in transitions between tables and graphs, whereas they had most difficulties with transitions that included a formula. More detailed information on students' reasoning and their use of strategies and frequent errors was obtained in the follow-up study by the same authors (Ceuppens *et al.*, 2018).

We have described here only the most frequently used multiple-choice instruments for measuring different aspects of understanding graphs. Although multiple-choice instruments are a well-established and reliable method for assessing knowledge, researchers should be aware of their strengths and weaknesses. In particular, they should be aware of the challenges and limitations of using multiple-choice instruments to evaluate students' abilities to construct and interpret graphs. Berg and Smith reported numerous differences between the results of multiple-choice and free-response instruments

(Berg and Smith, 1994). Due to limited space, we will not report here on numerous open-ended instruments for assessing students 'understanding of graphs, but we emphasize that these two methods of evaluation are complementary and together give a more complete insight into students' ability to construct and interpret graphs.

20.6 INTERPLAY OF MATHEMATICS AND PHYSICS IN STUDENT UNDERSTANDING OF GRAPHS

Teachers and university faculty often expect students to apply their knowledge acquired in mathematics to other contexts, but it seems that it is not easy for students to rise above the context. Recognizing mathematical concepts in a different context requires a good understanding of the new context, physics, other sciences, or real- life problems, together with the needed mathematical knowledge (Potgieter *et al.*, 2008). Several studies on graphs (Bassok and Holyoak, 1989; Woolnough, 2000; Planinic *et al.*, 2012; Planinic *et al.*, 2013; Wemyss and van Kampen, 2013; Ivanjek *et al.*, 2016; and Ceuppens *et al.*, 2018) investigated the transfer of knowledge between mathematics and physics, with mostly negative results. It was found that secondary students, even those who do well in mathematics and physics, do not sufficiently connect the two domains, and some even find it inappropriate to transfer concepts from mathematics to physics (Woolnough, 2000). Students in university physics courses do not always possess the required mathematical knowledge to transfer, especially when advanced concepts, such as derivatives or integrals, are concerned (Nguyen and Rebello, 2011; and Christensen and Thompson, 2012). Some researchers have pointed to the problem of domain specificity of knowledge in physics that prevents transfer (Bassok and Holyoak, 1989), but the same problem is also present in mathematics (Michelsen, 2005). Michelsen (2005) suggests that the problem lies in the missing link between mathematics and physics and that the mathematical domain should be expanded by using examples from physics and from everyday life contexts in mathematics teaching to enhance transfer. Cognitive studies that have looked for the transfer of knowledge have also usually come up with mostly negative results (Bransford and Schwartz, 1999). Bransford and Schwartz (1999) have suggested shifting the view on transfer from the direct application perspective (successful application of knowledge acquired in one context to similar problems in different contexts) to a more dynamical view of preparation for future learning (PFL). In the PFL perspective, the focus is not only on what students can or cannot directly transfer and solve, but also on whether students are able to learn while they transfer. In this way, transfer can be considered a dynamic way of reconstructing knowledge (Cui, 2006) rather than just an application of the previously acquired knowledge in a different situation.

Not many studies have attempted to compare student reasoning about graphs in physics and other domains (Woolnough, 2000; Planinic *et al.*, 2012, 2013; Wemyss and van Kampen, 2013; and Ivanjek *et al.*, 2016). In the study of Woolnough, the same Australian secondary students were tested with a simple quiz when they were in Year 11 and Year 12. The quiz asked them to find the slope of the line graph related to Hooke's law for a spring and to provide an interpretation of it. Students were later interviewed about how they calculated the slope and how they felt units of slope should be handled. Students entering the Year 11 physics course were not very familiar with the concept of slope from mathematics. By Year 12 they mostly learned how to calculate slopes; however, very few students in both years assigned units to calculated slopes (3% and 17%, respectively, for Years 11 and 12) even though they were aware of the importance of units in physics. Through interviews, it was revealed that they opposed the idea of assigning units to a "mathematical" concept, such as slope, suggesting the reluctance to break the perceived conventions of the fields of mathematics and physics and to transfer knowledge from one to the other. The study of Wemyss and van Kampen (2013) found that the number of correct answers of Irish university students to a real-life problem involving water level vs time graph was much higher than the number of correct answers to the supposedly more familiar problem of determining the speed of an object from a distance-time graph. The reason for students' poorer performance on physics problems was attributed to their reliance on learned procedures in physics (e.g., use of formulas). This study also found evidence that students' mathematical knowledge of slope does not guarantee their success on problems involving slope in kinematics.

In a study on line-graph slope, Planinic *et al.* (2013) compared Croatian second- year high school students' understanding of the line graph slope in the domains of physics and mathematics. Also, 90 Croatian physics teachers were asked to rank the isomorphic questions according to their expected difficulty for students. They largely expected the physics questions to be easier for students because they were regarded as less abstract than the mathematics questions. However, it was found that students did better on mathematics than on physics questions. The main source of student difficulties with the concept of line graph slope in physics seemed like not to be their lack of mathematical knowledge, but rather their lack of ability to interpret the meaning of the line graph slope in a physics context. It was observed that the transfer of knowledge from mathematics to physics did not always occur, even though many students possessed the required mathematical knowledge. Also, the same student difficulty known as slope/height confusion was detected in both domains, but it occurred about twice as often in physics than in mathematics (Planinic *et al.*, 2012).

In their next study, Planinic *et al.* (2013) investigated the effect of the context on student understanding of graphs using eight sets of three isomorphic questions and compared item difficulties as well as student strategies in different domains (Ivanjek *et al.*, 2016). The three domains were mathematics without context, physics (kinematics) and mathematics in contexts other than physics, which did not require additional conceptual knowledge. Questions were administered to 385 first-year students at the Faculty of Science, University of Zagreb in Zagreb, Croatia and later also to 417 first-year students at the University of Vienna (Ivanjek *et al.*, 2017). Students who were either prospective physics or mathematics teachers or prospective physicists or mathematicians were tested before any formal instruction on graphs, so their knowledge on graphs came only from high-school mathematics and physics instruction. Five sets of questions referred to the concept of graph slope, and three to the concept of area under a graph. Four sets were in a multiple-choice format, and four sets were open-ended and explanations and/or necessary calculations were required, so that insight into the underlying student reasoning. Both sets of data (Croatian and Austrian students) were analyzed with the Rasch model and showed a good fit and the stability of the test construct. Interestingly, isomorphic questions from the same set usually differed quite significantly in difficulty, suggesting that they were perceived by students as different questions.

The results suggested that students interpret graphs best in mathematics without context. This can be attributed to the fact that mathematics questions seem to be more direct and require less processing of information and less conceptual understanding than parallel physics (kinematics) questions. Kinematics was found to be a difficult context for students, even though it was extensively covered in high school. The level of difficulty was statistically indistinguishable from the difficulty of other context problems with which students were far less familiar. It was concluded that context generally increases the difficulty of items. by increasing the cognitive demand on the students and acting as an additional barrier in the problem. The only exception may be very familiar contexts for students.

The main problems identified in the studies on graphs in physics and other contexts can be summarized in the following points (Planinic *et al.*, 2019):

1. Strategies that students use are often context-dependent and domain-specific. The preferred strategy for physics questions seems to be the use of physics formulas.
   Students' almost exclusive reliance on formulas in physics (and sometimes on those that are incorrect or inappropriate for the situation) seems to present an important obstacle for the development of better reasoning strategies in physics, and sometimes even presents an obstacle for the transfer of knowledge and reasoning developed in other domains to physics.
2. Students use a wider spectrum of strategies on context problems than on physics problems. Other context problems could be potentially useful in physics and mathematics teaching.
   Students' reasoning is often limited by the contexts and conventions of the disciplines in which their knowledge was acquired. In mathematics and physics, students seem to stick firmly to the conventions of those disciplines, but they seem to think more freely and creatively, and to transfer more of their knowledge in other contexts. Other context problems may therefore have the potential to expose and develop student reasoning more than standard domain-specific mathematics and physics questions and should be used more in both mathematics and physics teaching.
3. Students show similar difficulties with graph interpretation in all domains, but there are differences between their understanding of graph slope and area under a graph.
   In the teaching of kinematics, the interpretation of slope is usually much more emphasized than the interpretation of area under a graph, but developing student reasoning which leads towards the interpretation of area should also not be neglected. That could help develop and strengthen student understanding of the concept of a definite and indefinite integral in mathematics.

Although many physics teachers attribute student difficulties with graphs in physics to their presumed lack of mathematical knowledge, this must not always be true. Even if students have the needed mathematical knowledge, the transfer to a different domain is not guaranteed because of an additional step of interpretation of mathematical quantities in physics or other contexts. Some cases of transfer of some problem-solving strategies from physics to other contexts were found on the area items (e.g., dimensional analysis), indicating a possible PFL type of transfer (Ivanjek *et al.*, 2016).

Both mathematics and physics should work more on establishing links between common concepts and procedures in both disciplines and promote their integration in students' minds to a much larger extent than seems to be the case now.

20.7 EYE-TRACKING STUDIES

Measurement of eye movements is a method that is increasingly used in science education (Devetak and Glažar, 2021). So far, researchers mostly employed eye tracking to explore students' visual attention during problem solving. For a recent review of eye tracking in physics education research, we refer to the systematic literature review by Hahn and Klein (2022). Here we will mention PER studies using eye tracking for the investigation of students' understanding of graphs.

Early eye-tracking studies on understanding graphs included graphs from kinematics (Kekule, 2014) and other topics in physics such as spectrometry, gas laws, and electrical resistance (Thoms *et al.*, 2013, 2014). In these studies, qualitative analysis of eye-tracking data was used (Kekule, 2015a) and mostly examined students gaze plots and heat maps. The gaze data plot consists of a sequence of fixations (state when the eye remains fixated on a particular point over some time that is called fixation duration) and saccades (rapid movement of the eye from one fixation to another) and it can give some insight into students' strategies. Researchers can focus on one participant and make a comparison between different attended areas, or they can compare data from more students. Heat maps provide summarized results and use different colors to show how long students attended a particular area. Heat maps are mostly used to qualitatively compare the visual attention of different student groups and to indicate areas of interest for quantitative analysis. Qualitative analysis usually starts with the definition of areas of interest (AOIs) and subsequent evaluation of various eye-tracking measures, such as dwell time or visit duration (the length of time a person spends attending to a particular AOI), number of fixations, average fixation duration, number of revisits to an AOI, number of transitions between two AOIs, etc.

As mentioned above, the early eye-tracking studies were qualitative and they analyzed heat maps (Kekule, 2014) and gaze plots (Thoms *et al.*, 2013, 2014) of groups of students who performed best and worst on the test, and groups of students who answered a task correctly or incorrectly. Kekule used seven tasks mostly adopted from the TUGK test by Beichner (1994), while Thoms *et al.* (2013) used tasks from three different topics (resistances, gas laws, spectrometry) with three levels of difficulty (elementary questions required simple data extraction such as reading individual values, intermediate-level questions involved comparisons of individual values, and high-level questions required a deeper understanding of the data). In the follow-up quantitative study, Kekule reported no difference in the average fixation duration for the best and the worst performers, which suggested that both groups did not have problems with the perceptual extraction of information (Kekule, 2015b). However, they differed in the distribution of attention to different parts of the graphs. For example, the best performing students attended different slopes to determine velocity from a position-time graph, whereas the worst performing students paid more attention to options representing common alternative conceptions.

In eye-tracking studies, the eye movements of experts and non-experts (novices) in the field are often compared. Such studies provide insight into experts' and novices' strategies in problem solving and could have important implications for teaching novices the perceptual and conceptual strategies of experts. Thus, Susac *et al.* compared physics and non-physics (psychology) students' understanding of graph slope and area under a graph (Susac *et al.*, 2018). The results confirmed the previous finding that the area under a graph is a difficult concept (Planinic *et al.*, 2013) and suggested that it is unlikely to be developed without formal teaching and learning. Psychology students scored much better on qualitative questions than on quantitative questions, whereas physics students solved them equally well, thus suggesting that studying physics helps students to quantitatively express relationships between quantities. Besides physics questions, students solved isomorphic questions with graphs related to prices (finance questions), which required the same mathematical procedure as physics (kinematics) questions. As expected, eye-tracking measures indicated that the physics context was easier for physics students as they had shorter total and axes dwell times for physics than the finance questions. However, the results provided indirect evidence for the transfer of knowledge from physics to finance because physics students solved the finance questions that were novel for them relatively well, but they used a similar procedure as in physics questions. Physics students strongly relied on the use of formulas, while psychology students mostly used common-sense strategies.

In a replication study comparing first-year physics and economics students' understanding of graphs, Klein *et al.* (Klein *et al.*, 2019) reported mostly similar findings as Susac *et al.* (Susac *et al.*, 2018). They found that attention to concept-specific areas of interest within the graphs discriminates the correct from the incorrect performers. Moreover, analysis of the confidence level of the two student groups revealed that physics students were better at judging their own performance than economics students. A postreplication study with a pretest-posttest design showed specific differences in the development of graph understanding over the first semester for physics and economics students (Brückner *et al.*, 2020). For example, all students had a shorter dwell time on the posttester than on the pretest, thus indicating that previous experience and familiarity with tasks can facilitate their comprehension and answer. In addition, it was found that students rated their performance less accurately on the posttester than on the pretest. Additional analysis of the data from economics students with a novel approach using epistemic network analysis revealed that incorrect solvers often had problems transferring textual information into graphical information and relied more on partly irrelevant parts of a graph (Brückner *et al.*, 2020).

A recent study in which data from the two previous studies (Susac *et al.*, 2018; and Klein *et al.*, 2021a) were aggregated and reanalyzed showed the differences in visual attention between isomorphic questions in physics and finance (Klein *et al.*, 2021b). When physics students solved questions in an unfamiliar context (finance), they needed more time to develop the strategy than in a familiar context (physics). They also spent more time attending to irrelevant parts of the graph in the finance context than in the physics context. These results further confirmed the important role of context in working with graphs.

In the previous chapter, we mentioned that the TUG-K is the most well-known instrument for measuring the understanding of graphs (Beichner, 1994). In 2020, Klein *et al.* reported the results of a measurement of eye movements of high school students while solving the TUG-K (Klein *et al.*, 2020). They found that students who correctly answered a question spent most time attending the correct option, while students who incorrectly answered the question most attended strong distractors, which represent common alternative conceptions. These results corroborate previous findings that when solving multiple-choice science problems, students paid more attention to chosen options than the rejected alternatives (Tsai *et al.*, 2012). Further cluster analysis of students' responses on the TUG-K using the transition metrics revealed three groups of items corresponding to predefined objectives (Klein *et al.*, 2021). This indicated that eye tracking can be useful in the evaluation and validation of test items in the process of instrument development.

As previously mentioned, instruments that measure representational competence in physics often contain items with graphs, as is the case with the Representational Variant of the Force Concept Inventory (R-FCI) (Nieminen *et al.*, 2010). Viiri *et al.* and Kekule and Viiri used some items from the R-FCI and found that the graph representation was easier for students than motion map, which indicates the position of an object at different times (Viiri *et al.*, 2017; and Kekule and Viiri, 2019). In another study, the authors explored how strategies depended on whether students preferred text or graph representations of the multiple-choice alternatives (Viiri *et al.*, 2020). They found that students who preferred a graph representation looked more at the graph options than the text alternatives, and correspondingly students who preferred a textual representation attended more textual options. A recent study on scientific argumentation with multiple representations also used problems with graphs (Wu and Liu, 2021). The results showed that the high-prior-knowledge group had better argumentation performance and more transitions between representations compared with the low-prior-knowledge group.

Some of the early PER studies that used eye tracking also contained tasks with graphs (Madsen *et al.*, 2012, 2013) and found that short visual cues can improve students' reasoning on introductory physics problems containing graphs (Madsen *et al.*, 2013). On the other hand, task-irrelevant data can impair processing during graph reading tasks (Strobel *et al.*, 2018). Skrabankova *et al.* (2020) showed that analysis of eye-tracking data during reading graphs can help in more accurate allocation of students to different groups according to their abilities.

20.8 CONCLUSION

Student understanding of graphs has been extensively investigated in the last three decades because graphs are an important representation of physics concepts and their functional relationships, crucial for developing students' understanding of physics and also important for their general scientific literacy. The research focused mostly on the interpretation of graphs and somewhat less on the construction of graphs. The findings formed the basis for the development of several diagnostic instruments on graphs, which can help teachers and university faculty to get an insight into the prevalence of student difficulties at different levels of physics learning.

The most common difficulties with graph interpretation are the interval/point confusion, slope/height confusion, iconic interpretations and difficulties interpreting the area under the graph. They can be expected at nearly all learning levels. Constructing a graph is also not an easy task for many students, and although some technical difficulties in this area can be circumvented with the use of computers (e.g., scaling of axes), some important difficulties remain (e.g., deciding on the variables on the graph).

Eye-tracking studies provide a new way of studying students' approach to problems containing graphs. The study of students' visual attention during problem solving can provide important insight into their strategies and the cognitive load of problems.

Very often, it is assumed that the difficulties with graphs in physics stem from students' inadequate mathematical knowledge. Studies that compared student difficulties with graphs in these two contexts suggest that this assumption is not true. Students in general seem to interpret graphs best in mathematics without any additional context. Pure mathematics questions are more direct and require less processing of information and less conceptual understanding than questions that are embedded in physics or real-life contexts. However, even when students have the necessary mathematical knowledge, the transfer of that knowledge into physics is still not guaranteed. Such transfer should be actively promoted in physics teaching and stronger links between mathematics and physics knowledge should be established.

An important implication of the research on graphs for physics teaching is that physics instructors should not only train students in graph procedures but also work more on developing students' conceptual understanding and reasoning about graphs.

REFERENCES

Araujo, I. S. *et al.*, Comput. Educ. **50**, 1128–1140 (2008).

Arteaga, P. *et al.*, *Statistical Graphs Complexity And Reading Levels: A Study With Prospective Teachers*, edited by M. Bassok and K. J. Holyoak (Statistique Et Enseignement, 1989). Correction to Bassok and Holyoak, J. Exp. Psychol. Learn. Memory Cognit. **15**(5), 867 (1989).

Bassok, M., and Holyoak, K. J., J. Exper. Psych.: Learn. Mem. Cog. **15**, 867–867 (1989).

Beichner, R. J., J. Res. Sci. Teach. **27**(8), 803–815 (1990).

Beichner, R. J., Am. J. Phys. **62**(8), 750–762 (1994).

Beichner, R. J., Am. J. Phys. **64**(10), 1272–1277 (1996).

Bektasli, B. and White, A. L., Eur. J. Educ. Res. **48**, 1–19 (2012).

Bell, A. and Janvier, C., Learn. Math. **2**, 34–42 (1981).

Berg, C. A. and Boote, S., Int. J. Sci. Math. Educ. **15**(1), 19–38 (2017).

Berg, C. A. and Phillips, D. G., J. Res. Sci. Teach. **31**(4), 323–344 (1994).

Berg, C. A. and Smith, P., Sci. Educ. **78**(6), 527–554 (1994).

Bollen, L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(1), 010108 (2016).

Bransford, J. D. and Schwartz, D. L., Rev. Res. Educ. **24**, 61 (1999).

Brasell, H. M. and Rowe, M. B., School Sci. Math. **93**(2), 63–70 (1993).

Brückner, S. *et al.*, Sensors (Basel, Switzerland) **20**(23), 6908 (2020a).

Brückner, S. *et al.*, Front. Psychol. **11**, 2090 (2020b).

Cagande, J. L. L. and Jugar, R. R., Issues Educ. Res. **28**, 288–307 (2018).

Carli, M. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010111 (2020).

Carpenter, P. A. and Shah, P., J. Exp. Psychol. Appl. **4**, 75–100 (1998).

Ceuppens, S. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(2), 020105 (2018).

Cheng, P. C.-H. *et al.*, Artif. Intell. Rev. **15**(1/2), 79–94 (2001).

Christensen, W. M. and Thompson, J. R., Phys. Rev. Spec. Top.: Phys. Educ. Res. **8**(2), 023101 (2012).

Cui, L., *Assessing College Students' Retention and Transfer from Calculus to Physics* (Kansas State University, 2006).

Curcio, F. R., J. Res. Math. Educ. **18**(5), 382 (1987).

Devetak, I. and Glažar, S. A., *Applying Bio-Measurements Methodologies in Science Education Research* (Springer International Publishing, Cham, 2021).

Dominguez, A. *et al.*, EURASIA J. Math. Sci. Technol. Educ. **13**(10), 6507 (2017).

Dossey, J. A. *et al.*, Can students do mathematical problem solving? Results from constructed-response questions in NAEP's 1992 mathematics assessment. Nation's report card no. 23-FR01 (U.S. Dept. of Education, Office of Educational Research and Improvement, Washington, 1993).

Dreyfus, T. and Eisenberg, T., in *Proceedings of the 11th International Conference of PME* (IGPME, Montreal, 1987), Vol. l, pp. 190–196.

Forster, P. A., Res. Sci. Educ. **34**(3), 239–265 (2004).

Gal, I., Int. Stat. Rev. **70**(1), 1–25 (2002).

Gattis, M. and Holyoak, K. J., Learn. Memory Cognit. **22**(1), 231–239 (1996).

Geyer, M.-A., dissertation (Technische Universität Dresden, 2019).

Glazer, N., Stud. Sci. Educ. **47**(2), 183–210 (2011).

Goldenberg, E. P., "Mathematical, technical, and pedagogical challenges in the graphical representation of functions," [Final] Technical Report (1988), pp. 88–84.

Hahn, L. and Klein, P., Phys. Rev. Phys. Educ. Res. **18**(1), 013102 (2022).

Hattikudur, S. *et al.*, School Sci. Math. **112**, 230–240 (2012).

Hill, M. and Sharma, M. D., EURASIA J. Math. Sci. Technol. Educ. **11**(6), 1633 (2015).

Hill, M. *et al.*, Int. J. Innovation Sci. Math. Educ. **22**(6), 22–42 (2014).

Ivanjek, L. *et al.*, *Contributions From Science Education Research. Cognitive and Affective Aspects in Science Education Research*, edited by K. Hahl *et al.* (Springer International Publishing, Cham, 2017), Vol. 3, pp. 167–178.

Ivanjek, L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(1), 010106 (2016).

Jacobbe, T. and Horton, R. M., Stat. Educ. Res. J. **9**, 27–45 (2010).

Kekule, M., (Science Education research Group as Eastern Mediterranean University, Famagusta, 2014), pp. 108–117. European Journal of Science and Mathematics Education.

Kekule, M., (*International Conference on Contemporary Issues in Education* (GLOBE-EDU, 2015a), pp. 104–111.

Kekule, M., (*International Conference on Contemporary Issues in Education* (GLOBE-EDU, 2015b), pp. 126–134.

Kekule, M. and Viiri, J., Scientia in Educatione **9**(2), 117–130 (2019).

Kerslake, D., Math. School, **6**, 22–25 (1977).

Kerslake, D., *Students's Understanding of Mathematics*, edited by K. Hart (John Murray, London, 1981), pp. 11–16.

Klein, P. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(1), 013102 (2021a).

Klein, P. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020116 (2019).

Klein, P. *et al.*, *Applying Bio-Measurements Methodologies in Science Education Research*, edited by I. Devetak and S. A. Glažar (Springer International Publishing, Cham, 2021b), pp. 243–260.

Klein, P. *et al.*, Eur. J. Phys. **41**(2), 025701 (2020).

Klein, P. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(1), 010132 (2017).

Knight, S. A., *Understanding Graphs* (Blackie, Glasgow, 1966).

Kosslyn, S. M., *Elements of Graph Design* (Freeman, New York, 1994).

Kozhevnikov, M. *et al.*, Cognit. Sci. **31**(4), 549–579 (2007).

Kozhevnikov, M. and Thornton, R., J. Sci. Educ. Technol. **15**(1), 111–132 (2006).

Lachmayer, S., Entwicklung und Überprüfung eines Strukturmodells der Diagrammkompetenz für den Biologieunterricht. IPN - Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik an der Universität Kiel (2008).

Lai, K. *et al.*, J. Sci. Educ. Technol. **25**(4), 665–681 (2016).

Leinhardt, G. *et al.*, Rev. Educ. Res. **60**(1), 1–64 (1990).

Lewandowsky, S. and Spence, I., J. Am. Stat. Assoc. **84**, 682–688 (1989).

Lichtenberger, A. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(1), 010115 (2017).

Lohman, D. F., *Human Abilities: Their Nature and Measurement*, edited by I. Dennis and P. Tapsfield (Lawrence Erlbaum Associates, Inc., 1996), pp. 97–116.

Madsen, A. M. *et al.*, Phys. Rev. Spec. Top.: Phys. Educ. Res. **8**(1), 010122 (2012).

Madsen, A. *et al.*, Phys. Rev. Special Top. Phys. Educ. Res. **9**(2), 020104 (2013).

Maries, A. and Singh, C., Phys. Rev. Spec. Top. Phys. Educ. Res. **9**(2), 020120 (2013).

Markovits, Z. *et al.*, *Proceedings of the Seventh International Conference of the International Group for the Psychology of Mathematics Education*, edited by R. Hershkowitz (Weizmann Institute of Science, Rehovot, 1983), pp. 271–277.

McDermott, L. C. *et al.*, Am. J. Phys. **55**(6), 503–513 (1987).

McKenzie, D. L. and Padilla, M. J., J. Res. Sci. Teach. **23**(17), 571–579 (1986).

Mevarech, Z. R. and Kramarsky, B., Educ. Stud. Math. **32**(3), 229–263 (1997).

Michelsen, C., *Proceedings of the 1st International Symposium of Mathematics and its Connections to the Arts and Sciences*, edited by A. Beckmann *et al.* (University of Education Schwäbisch Gmünd, Germany, 2005), pp. 201–214.

Mokros, J. R. and Tinker, R. F., J. Res. Sci. Teach. **24**(4), 369–383 (1987).

Nguyen, D.-H. and Rebello, N. S., Phys. Rev. Spec. Top. Phys. Educ. Res. **7**(1), 010112 (2011).

Nieminen, P. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **6**(2), 020109 (2010).

Nixon, R. S. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(1), 010104 (2016).

Padilla, M. J. *et al.*, School Sci. Math. **86**(1), 20–26 (1986).

Patahuddin, S. M. and Lowrie, T., Int. J. Sci. Math. Educ. **17**(4), 781–800 (2019).

Pinker, S., *Artificial Intelligence and the Future of Testing*, edited by R. Freedle (Lawrence Erlbaum Associates, Inc., 1990), pp. 73–126.

Planinic, M. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **9**(2), 020103 (2013).

Planinic, M. *et al.*, Int. J. Sci. Math. Educ. **10**(6), 1393–1414 (2012).

Planinic, M. *et al.*, *Mathematics in Physics Education*, edited by G. Pospiech *et al.* (Springer International Publishing, Cham, 2019), pp. 233–246.

Pospiech, G. *et al.*, J. Phys.: Conf. Ser. **1287**(1), 012014 (2019).

Pospiech, G. *et al.* “Making meaning of graphical representations in beginners’ physics lessons,”in *ESERA 2011 Conference. Part 6: Discourse and Argumentation in Science Education* (ESERA, 2011), pp. 65–71.

Potgieter, M. *et al.*, J. Res. Sci. Teach. **45**(2), 197–218 (2008).

Roth, W.-M. and McGinn, M. K., Sci. Educ. **81**(1), 91–106 (1997).

Schiano, D. J. and Tversky, B., Memory Cognit. **20**(1), 12–20 (1992).

Shah, P. and Freedman, E. G., Top. Cognit. Sci. **3**(3), 560–578 (2011).

Shah, P. *et al.*, *The Cambridge Handbook of Visuospatial Thinking*, edited by P. Shah and A. Miyake (Cambridge University Press, 2012), pp. 426–476.

Shah, P. and Hoeffner, J., Educ. Psychol. Rev. **14**(1), 47–69 (2002).

Skrabankova, J. *et al.*, J. Baltic Sci. Educ. **19**(2), 298–316 (2020).

Strobel, B. *et al.*, Learn. Instr. **55**, 139–147 (2018).

Susac, A. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(2), 020109 (2018).

Thoms, L. J. *et al.*, PhyDid B-Didaktik Der Physik-Beiträge Zur DPG-Frühjahrstagung (published online 2013).

Thoms, L.-J. *et al.*, *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht. Gesellschaft für Didaktik der Chemie und Physik - Jahrestagung*, edited by S. Bernholt (2014), Vol. 34, pp. 513–515.

Torres, T. S. and Alarcon, H., Lat. Am. J. Phys. Educ. **6**(Suppl. I), 285–289 (2012).

Tsai, M.-J. *et al.*, Comput. Educ. **58**(1), 375–385 (2012).

Van den Eynde, S. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020113 (2019).

Viiri, J. *et al.*, “Eyetracking the effects of representation on students’ problem solving approaches,” in *Proceedings of the Annual FMSERA Symposium* (Finnish Mathematics and Science Education Research Association, 2017).

Viiri, J. *et al.*, *Challenges in Physics Education. Research and Innovation in Physics Education: Two Sides of the Same Coin*, edited by J. Guisasola and K. Zuza (Springer International Publishing, Cham, 2020), pp. 145–154.

Wavering, M. J., J. Res. Sci. Teach. **26**(5), 373–379 (1989).

Wemyss, T. and van Kampen, P., Phys. Rev. Spec. Top. Phys. Educ. Res. **9**(1), 010107 (2013).

Woolnough, J., Res. Sci. Educ. **30**(3), 259–267 (2000).

Wu, C.-J. and Liu, C.-Y., Phys. Rev. Phys. Educ. Res. **17**(1), 010125 (2021).

Zavala, G. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 020111 (2017).

Zawojewski, J. S. and Heckman, D. J., *Results From the Sixth Mathematics Assessment of the National Assessment of Educational Progress*, edited by P. A. Kenney and E. A. Silver (National Council of Teachers of Mathematics, Reston, VA, 1997).

CHAPTER

21 VISUALIZATION AND MATHEMATIZATION: HOW DIGITAL TOOLS PROVIDE ACCESS TO FORMAL PHYSICS IDEAS

Elias Euler, Lorena Solvang, Bor Gregorcic, and Jesper Haglund

Euler, E., Solvang, L., Gregorcic, B., and Haglund, J., "Visualization and mathematization: How digital tools provide access to formal physics ideas," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 21-1–21-28.

21.1 INTRODUCTION

Visual depictions in physics rarely purely *depict* objects and phenomena as such. Instead, visual depictions also imbue the details of physics phenomena with formal mathematical structures and operations (Lynch, 1988). Consider, for example, some common visual depictions involved in the teaching and learning of physics, each of which necessitates the inclusion of mathematical formalism: free-body diagrams entail schematizing objects as points and indexing relevant variables (Rosengrant *et al.*, 2009); kinematic graphs involve quantifying physical properties of the system at hand and defining labelled axes; the renderings of computer-based physics simulations are all in some way "built up from the formal mathematical relationships in [their] source code" (Euler and Gregorcic, 2019, p. 361).

Herein lies what can be called the interrelation of *visualization* and *mathematization* in physics education: to visually depict generally entails the incorporation of mathematical formalism and often vice-versa. However, while the topics of visualization and mathematization have often featured separately in physics education research (PER) literature, it is uncommon for the two topics to be

https://doi.org/10.1063/9780735425514_021

discussed together in a manner that enables an examination of their apparent interplay.[1] In this chapter, we will synthesize the physics education research work done at the intersection of visualization and mathematization in physics, specifically when visualization and mathematization are mediated by digital technologies. Before doing so, though, it is important to first disentangle the numerous meanings that both visualization and mathematization have accrued in the fields of science/mathematics education research and cognitive science over the past century and a half.

Since the late 1980s[2], it has been most common in physics/science to see the term *visualization* used to refer to the visual portrayal of data achieved via digital technologies (e.g., Tversky, 2005; and Kohnle *et al.*, 2015). In education research fields such as PER, however, one finds two main senses of the term visualization related to internal and external processes. In one sense, visualization is used to refer to the mental imaging of information "in the mind's eye" (Galton, 1883; Ganguly, 1995; and Phillips *et al.*, 2010): to *internally* visualize is to privately picture something in a manner that it is not immediately accessible to others. With internal visualization, other terms are often used, such as mental imagery (e.g., Reiner, 1998; Clement, 2008; and Stephens and Clement, 2010) or mental simulation (e.g., Monaghan and Clement, 1999; and Alibali and Nathan, 2012). In a second sense, visualization is used to denote the visual display of information (Tufte, 1983): to *externally* visualize is to take something that is less visually accessible, such as an abstract idea or tabulated dataset, and render it in a publicly visible format. Physics education researchers have often foregone the term "external visualization" for related words like representation to refer to the products of such externalization processes (e.g., Van Heuvelen, 1991; Kohl and Finkelstein, 2005; and Fredlund *et al.*, 2015).

The connection between these two, internal and external, uses of the term runs deeper than a mere semantic association. The two processes are evidently related to one another when one considers how doing/teaching/learning physics relies on back-and-forth communication via (external) public depictions and sense-making via (internal) mental imagery (Chen and Gladding, 2014; Euler *et al.*, 2019; and Samuelsson, 2020; see also, work like Schnotz and Kürschner, 2008). Occasionally both the internal and external senses of the visualization are used by PER authors within the same publication, with some papers tacitly using *visualization* as something external a computer produces and *visualize* as something a person does mentally (McKagan *et al.*, 2008), and other papers clearly label visualizations as internal or external where appropriate (Lingefjärd and Ghosh, 2016; and Mešić *et al.*, 2016).

Mathematization, on the other hand, is in one sense used to describe the broader introduction of mathematical methods to a discipline or theory (Kuhn, 1970; Gingras, 2001; and Uhden *et al.*, 2012): to

---

[1] The dearth of literature on the interplay of visualization and mathematization is in part due to how much of the relevant research on these topics avoids these specific terms. This is a semantic problem resulting from the manifold ways visualization and mathematization in physics are discussed. However, even when other terminology is used, it is rarely made explicit how visualization and mathematization in physics are interrelated and what this interrelation means for physics education.

[2] See *Visualization of Scientific Computing* (McCormick *et al.* (ed), 1991) and the intentional co-opting of the term "visualization" therein (Wolfe, 1988; Defanti and Brown, 1991; and Phillips *et al.*, 2010).

historically mathematize some otherwise nonmathematical domain is to bring mathematics to bear on the topics within that domain. However, physicists often use mathematization in a second, narrower sense, whereby the term refers to the localized translation of physical phenomena into mathematical structures and formulas (Niss, 2017): to *interpretatively* mathematize some phenomenon is to describe and examine that phenomenon by way of mathematics. In line with this second sense of the word, Brahmia (2014) writes that "to mathematize in physics means going back and forth between the physical world and the symbolic world" (p. 11). As such, mathematization in the interpretative sense is closely related to the notion of mathematical *modeling*—a core practice of physics (Hestenes, 1992; Clement, 2008; Redish and Bing, 2009; and Uhden *et al.*, 2012).

This chapter deals with the latter sense for both visualization and mathematization—i.e., external visualization and interpretative mathematization. More specifically, in keeping with the widely adopted usage of visualization as the external renderings of computers, we focus on how digital technologies mediate external visualization and interpretative mathematization in physics education. We synthesize the relevant PER work done on digital depictions in physics and examine how those depictions support physics students' translation of physical phenomena into mathematical formalism. For the sake of clarity, we refer to these digital depictions as *visualization tools*.

An important detail is that in this chapter, visualizations are interpreted as representations of aspects of the physical world distinct from the phenomena they are meant to represent. This chapter thereby connects to the general issue of the role of representations in physics education—taken up further in Chapter 4.7 of this handbook—yet focuses on the dynamic, interactive visualizations provided through digital technologies. In this way, we stress that many interesting and relevant types of visualizations are left out of our literature synthesis, such as static, non-digital drawings on paper/whiteboards (e.g., Wenning, 2005) or non-interactive simulations and video clips of physical phenomena (as can be used during lectures, for instance; see Wieman *et al.*, 2010).

21.2 THEORIES OF VISUALIZATION/MATHEMATIZATION AND OUR APPROACH TO SYNTHESIS

Existing PER work (and relevant work from adjacent fields) on visualizations and mathematization has overwhelmingly focused on one or the other of these topics and has often done so without explicit use of the terminology "visualization" or "mathematization." This means that, unlike many of the topics taken up in the other chapters of this handbook, the interplay of visualization and mathematization in physics cannot be readily synthesized into any small number of clearly defined citation lineages or be epitomized by any one community of researchers convergently building upon one another's work. Instead, our synthesis in this chapter has involved bringing together a diverse collection of research efforts that all generally relate to some partial aspects of visualization and mathematization in physics from a variety of theoretical perspectives. In this section, we highlight some of the notable ways that researchers have come to view either visualization or mathematization, and ultimately discuss the work which most directly underpins our synthesis of the interplay of the two topics via visualization tools.

21.2.1 Theories of visualization in physics

External visualizations and their role in the doing/teaching/learning of science have increasingly occupied the attention of historians of science and education scholars. Latour (1990) famously asserted that the drive to seek out and depict the world has been crucial for scientific progress, while several physicists and education scholars early to implement computers in physics teaching (e.g., Schwartz and Taylor, 1968; and White, 1984) identified the power that depictions could hold for the process of learning physics. As of writing, there are many well-established educational theories relevant to visualization in physics, with the most commonly used emphasizing either the limits to students' cognitive processing, the benefits of combining multiple representations, the influence of socio-cultural processes, and/or the role of sensory-motor experiences.

Cognitive theories such as cognitive load theory (Sweller, 1988, 1989; and Chandler and Sweller, 1991), the cognitive theory of multimedia learning (Mayer, 1997, 2014), and the integrated model of text and picture comprehension (Schnotz, 2014) have been used to explore the extent to which certain visualization tools may overwhelm the functional limit of physics students' cognitive capacity (Lee *et al.*, 2004; Wu *et al.*, 2015; Zu *et al.*, 2018; Strzys *et al.*, 2019; and Thees *et al.*, 2020). With such cognitive theories, scholars will often infer that ineffective visualization tools fail by, for example, splitting students' attention (Chandler and Sweller, 1992), requiring too much direct problem solving instead of worked examples (Sweller, 1988), or generally overwhelming students' cognitive load via extraneous material (Mayer and Moreno, 2003). For more on cognitive perspectives in PER, see Sec. II of this handbook.

Ainsworth (1999, 2006) proposed a theory of multiple (external) representations for designing and examining the interplay of several visualizations used simultaneously with one another. While this specific theory and the DeFT (Design, Functions, Tasks) framework associated with it (Ainsworth, 2006) have not been explicitly featured in many PER publications to date, the general topic of multiple representations in physics has received significant and sustained attention among PER scholars (e.g., Van Heuvelen and Zou, 2001; van der Meij and de Jong, 2006; and Treagust *et al.*, 2017). Some of the more recent research on this idea specifically emphasizes the role of visualization tools in students' movement between representations (Volkwyn *et al.*, 2019; and Svensson *et al.*, 2020). For more information on multiple representations in PER, see Chapter 4.2 of this handbook.

There are also theories such as (cultural historical) activity theory (Roth and Lee, 2007) and social semiotics (Hodge and Kress, 1988; and Airey and Linder, 2017) that have been used in PER to foreground the social-cultural aspects of visualizations and their role in doing/teaching/learning physics (Eriksson, 2014; Gregorcic, 2015a; and Svensson *et al.*, 2020). Research from authors such as Hollan *et al.* (2000) has blurred the lines between cognitive and socio-cultural theories through their work on distributed cognition and technology, though distributed cognition has so far made little headway into the work of PER scholars studying visualization tools.

Various other research efforts, such as those associated with the theory of grounded cognition (Barsalou, 2008) and the field of human-computer interaction (Card *et al.*, 1983), foreground the sensory-motor experiences of students during physics learning through visualizations (Chen and Gladding, 2014). Such works often comprises examinations of virtual reality (VR) interfaces and physics learning, with Dede *et al.* (1999); and Whitelock *et al.*'s (1996) research on immersion being core theoretical waypoints for research delving into such topics (see also, Chapter 4.3 of this handbook). Visualization-related PER work using theories related to sensory-motor experiences tends to emphasize the importance of building on students' embodied intuitions through visualization design (Chen and Gladding, 2014).

21.2.2 Theories of mathematization in physics

Theories related to mathematization in physics are fewer in number than those related to visualization, though the topic has similarly garnered growing attention among PER scholars, especially in the last two decades. The major strands of the theory related to mathematization in physics (see Niss, 2017) focus on the form of mathematics that physics students encounter, their epistemologies, and/or mathematical modeling.

Building on cognitive traditions, Sherin (1996, 2001) described the so called "symbolic forms" that physics equations can take in students' minds. This perspective advocates acknowledging the general template and conceptual importance of physics equations to facilitate more robust connections between mathematical problem solving and conceptual understanding (e.g., Dreyfus *et al.*, 2017; and Ryan *et al.*, 2018). A hallmark of this school of thought around mathematization is attention to how students make sense of physics equations rather than how they translate between physical phenomena and mathematics. Thus, compared to some of the other perspectives on mathematization in physics education, symbolic forms are arguably less about the back-and-forth between the physical and mathematical, as Brahmia (2014) describes, and more about a type of sensemaking between mathematical and conceptual domains.

Other theoretical perspectives on mathematization, such as the epistemic perspective put forth by Collins and Williams (1993), have been used in PER to emphasize the types of mindset students inhabit as they go about solving physics problems (Hammer *et al.*, 2005; Black *et al.*, 2007; Tuminaro and Redish, 2007; and Bing and Redish, 2009). Though generally also based on cognitive principles like Sherin's (1996, 2001) work with symbolic forms, these epistemological theories depart by clearly highlighting some of the difficulties students face in translating from physical scenarios to mathematics and vice-versa (Steinberg *et al.*, 1997; and Redish *et al.*, 1998).

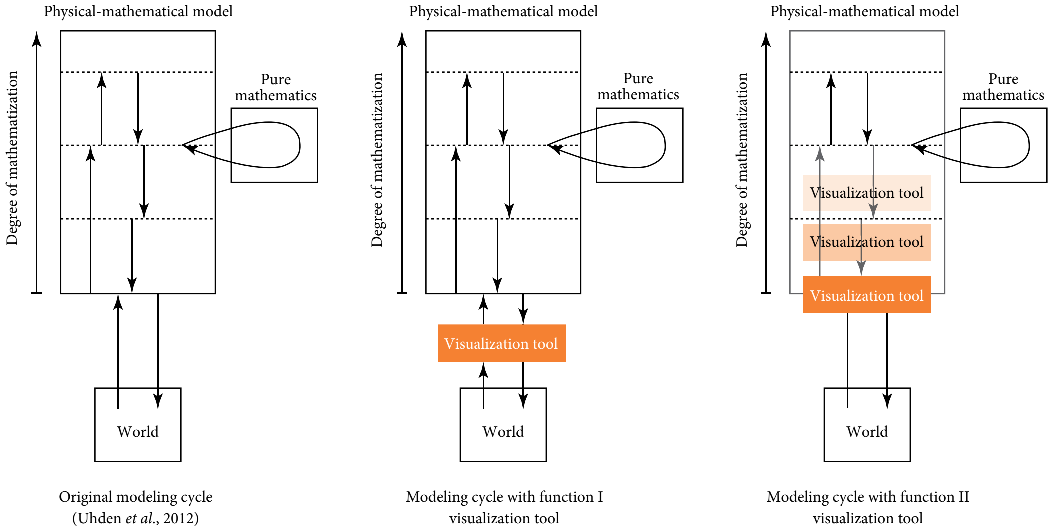
Implicit or explicit in nearly all these mathematization theories is the notion that physics entails some cycles of mathematical modeling. Hestenes (1987, 1992) described the process of performing physics as a "modeling game," highlighting how physicists move between the physical domain and a formal (mathematical) domain in a two-part cycle via a set of disciplinary rules. Redish (Redish, 2005; and Redish and Bing, 2009) proposed a four-part cycle for mathematical modeling in physics where an original physical context is first mapped onto mathematics, this mathematics is processed into mathematics of another form (through solving or derivation), the processed mathematics is decoded into a physical interpretation, and finally the physical interpretation is evaluated based on how well it describes the original physical situation.

Many such "models of modeling" abound in the physics and mathematics education research literature (e.g., Blum and Borromeo, 2009; see also, the chapters in this section of the handbook). However, Uhden *et al.* (2012) add important nuance in this area by highlighting how not all acts of translating into mathematics are equally mathematized. That is, there are *degrees of mathematization* when modeling in physics worthy of disambiguation. Though it is not always straightforward to judge what is more or less mathematized, one example given by the authors concerns velocity: the phrase "distance over time" is less mathematized than $v = \frac{\Delta x}{\Delta t}$, which is in turn less mathematized than $\vec{v} = \frac{d\vec{r}}{dt}$ (Uhden *et al.*, 2012, p. 498).

21.2.3 Bringing theories of visualization and mathematization together

In this chapter, we build on a perspective published by diSessa (1988) that not only deals with the interplay of visualization and mathematization for physics but also explicitly proposes a mechanism for how computer-based visualization tools provide an advantage for students during mathematization. diSessa (1988), a scholar in education research who was early to examine the potentially-transformative role of computers in physics education (Papert *et al.*, 1979; and diSessa, 1980, 1982), suggested that visualization tools could act as *semi-formalisms* for students as they learned physics—a term he used to emphasize how computer-based environments were "manipulable systems that can serve as general and precise formalism but which retain for students a sense of familiarity and evident controllability" (diSessa, 1988, p. 64). For diSessa, visualization tools could assist students in mathematization in physics by rendering the mathematical formalism through manipulable systems, thereby portraying mathematical details in a way that more closely resembled the physical world.

We have previously built on the notion of semi-formalisms in physics learning (Euler and Gregorcic, 2018) by combining diSessa's (1988) perspective on visualization tools with the modeling framework of Hestenes (1992). In that work, we propose that semi-formal visualization tools can aid mathematization by acting as a type of intermediate "steppingstone" between the physical world and the formal (mathematical) world. However, for the purposes of our explicit attention on mathematization in this chapter, we further refine our extension of diSessa's theory by integrating semi-formal visualization tools into the theory of modeling with degrees of mathematization by Uhden *et al.* (2012). In this



Original modeling cycle (Uhden *et al.*, 2012)

Modeling cycle with function I visualization tool

Modeling cycle with function II visualization tool

FIG. 21.1

Mathematization/modeling cycle of physics by Uhden *et al.* (2012), recreated (left) and modified to show the figurative placement of visualization tools consistent with Function I (center) and Function II (right) from this chapter. Uhden *et al.*'s original schematic emphasizes the degrees of mathematization within a mathematized model as well as accounting for mathematical processing (via "pure mathematics") that moves neither "up" nor "down" on the mathematization axis.

updated approach, we highlight that visualization tools can functionally reside either between the physical world or a mathematized model (as before) or at some lower degree of mathematization within a mathematized model—keeping in mind that the goal for visualization tools in either position is to facilitate students' transitions back-and-forth between the physical and mathematical. Thus, we identify two distinct functions that visualization tools can serve in facilitating mathematization in physics (Fig. 21.1):

- **Function I**. *Bridging physical phenomena and formalism* by superimposing mathematical formalism onto physical phenomena and/or depicting formalism alongside physical phenomena.
- **Function II**. *Bridging idealized models of physical phenomena and formalism* by superimposing mathematical formalism onto models and/or depicting formalism alongside idealized models.

In both Function I and II lies a key benefit of semi-formalism visualization tools for mathematization in physics: they afford physics students mathematized yet experiential access points to physical phenomena, thereby providing opportunities for those students to leverage their intuitions for interpreting formal ideas. This includes the benefit of providing access to the mathematical models and abstract concepts common to physics that are often otherwise non-experienceable.

21.3 FUNCTION I: BRIDGING PHYSICAL PHENOMENA AND FORMAL REPRESENTATIONS

In this section and the one that follows, we review some of the research done around visualization tools that facilitate mathematization through Functions I and II. For Function I, we highlight the examples of MBL/probeware tools, video-analysis tools, and augmented/mixed reality. Each example illustrates how visualization tools can dynamically superimpose formal representations such as equations, vector arrows, and/or graphs on top of or alongside physical phenomena.

It is important to note that the visualization tools included throughout this chapter should not be binarily classified as *either* Function I or II tools—though the structure of our chapter implies such exclusive categorization. Many of the visualization tools described hereafter can exhibit either or both functions at times, especially depending on the manner and context in which they are used by teachers and students [see a similar argument for context-dependent judgments of educational technologies in Rieber (1996)].

21.3.1 Microcomputer-based laboratory (MBL)/probeware tools

The advent of the microcomputer revolutionized education in the 1970s onward, not least among educators using computer technologies to improve the teaching and learning of physics. Microcomputers were a true "harbinger of great computer power at low cost" (Solomon, 1986, p. 7), and this was quickly borne out in the construction of so-called microcomputer-based laboratory (MBL) tools—now often called probeware (Tinker, 2000). MBL/probeware tools involved various sensors connected to a computer, allowing students to collect data and generate time-series graphs of physics phenomena in real time (Tinker, 1981). The aim of MBL/probeware tools was to free students from the drudgery of having to arrange data in tables and draw graphs manually while collecting data (Tinker, 2000). Using computer-interfacing sensors, students can focus on interpreting graphs and connecting them to the studied phenomena—exemplifying the Function I aspect of visualization tools by facilitating students' back-and-forth movement between mathematized graphs and physical scenarios.

For example, Mokros and Tinker (1987) found that asking students to walk in front of an ultrasonic motion detector allowed them to connect their kinesthetic experiences of motion to formal distance graphs in real time. Thornton and Sokoloff (Thornton, 1987; and Thornton and Sokoloff, 1990) developed university physics curricula based on a similar use of MBL/probeware tools called *RealTime Physics* (Sokoloff *et al.*, 2007), finding that such approaches led to significant gains in the conceptual understanding of kinematics. In parallel, Priscilla Laws (1991) developed *Workshop Physics*, a curriculum where students develop inquiry skills through MBLs/probeware use in laboratory environments without traditional lectures.

21.3.1.1 MBL/probeware example: iOLab

In 2010, Mats Selen and colleagues started the development of the Interactive Online Laboratory system (iOLab), a portable, low-cost probeware tool equipped with a wide range of sensors and wirelessly connected to a computer for real-time graphing. The iOLab has since been integrated into introductory mechanics courses at Illinois (Ansell and Selen, 2018; and Ansell, 2020), promoting students' scientific skills and encouraging them to think like scientists. Recognizing the low cost and mobility of iOLab, Bodegom *et al.* (2019) used the device to adapt *RealTime Physics* for distance learning.

Volkwyn *et al.* designed teaching interventions and conducted qualitative video analyses of students' use of iOLab (Volkwyn *et al.*, 2018, 2020b). Volkwyn *et al.* (2019) identified that students learned about the Earth's magnetic field by turning the iOLab along different axes and thereby influencing the graphical display on the screen. In further analysis of the students' interaction, Volkwyn *et al.* (2020b) identified that the use of the iOLab device enabled the students to appreciate the movability of coordinate systems, in contrast to textbooks that often display coordinate systems in fixed ways. In supporting students' engagement with the mathematical tools of coordinate systems against the background of the physical system of the Earth's magnetic field, the use of the iOLab here can be seen to have facilitated mathematization in line with Function I.

21.3.2 Video-analysis tools (interactive video)

Video-analysis tools allow students to interact in various ways with segments of video to observe otherwise difficult-to-discern details, track and measure relevant variables, and test mathematical models against physical phenomena. Many video-analysis tools are linked to databases with prerecorded video sequences of real-word phenomena ready to be analyzed, but some also allow users to import video files from other sources. When analyzing a video, the user typically "marks" the positions of an object across several key video frames and establishes a scale and frame of reference. The software can then be instructed to calculate quantities such as displacement, velocity, acceleration, force, and energy, while also being able to construct graphs and report equations of best fit for these quantities.

Historically, capturing the video for analysis in the physics lab was not easy. A video source, such as a camera, cassette recorder, or videodisc player, had to be connected to a computer to digitize and store the incoming analog signal. To analyze and model data from digital video, the first PER scholars that implemented and studied video analysis—including Beichner (1996, 1999), Cadmus (1990); Escalada and Zollman (1997); and Laws and Pfister (1998)—developed their own so called "interactive video" software that allowed students to observe and/or measure relevant variables directly from the videos themselves. Early analysis of *Video Analyzer* and *Visual Space-Time* (Escalada and Zollman, 1997) showed that students felt the activities were effective in helping them learn the physics concepts related to reference frames, but there were no significant differences in final exam scores between the students who used the video-analysis tools and the students who did not.

In the time since the early videodisc software of the 1980s, it has become significantly easier to carry out video-based laboratories due to the availability of high-quality digital video cameras, fast-working conversion tools, and user-friendly software packages. Students can now easily capture the video of physical phenomena around them and analyze it immediately via video-analysis tools such as *Tracker*, *VideoPoint*, *Physics ToolKit*, and *Measurement-in-Motion* (Bryan, 2005). Research such as that from Wee *et al.* (2012); and Malgieri, Onorato, Macheretti *et al.* (2014) indicates that combining experimentation and video analysis can help students to build a link between concrete physical situations and the more formal representations (in accordance with Function I).

21.3.2.1 Video-analysis tool example: Tracker

Tracker is a free video-analysis and modeling tool built on the Open Source Physics engine.[3] Users of *Tracker* can have the motion of objects mathematized by marking the position of those objects at key video frames and allowing the software to generate lines of best fit. *Tracker* can also be used for video-based modeling (Brown, 2008), where students first define the theoretical equations for a physical phenomenon in the video and then this mathematical model is superimposed onto the video by the visualization tool frame-by-frame (Wee *et al.*, 2012; and Wee, Tan *et al.*, 2015). In this way, students can construct their own models of physical phenomena and compare them with physical phenomena. In line with Function I, Wee *et al.* (2012, 2015) suggest that since the values deduced from video analysis can be shown to be consistent with data collected in the physical world, *Tracker* allows students to connect abstract concepts and formulas with physical situations as they invent and improve models of physical phenomena.

The Tracker has been used to analyze phenomena such as non-thermal emission spectroscopy (Brown and Cox, 2009), the Beer–Lambert law (Onorato *et al.*, 2021), and the emission spectra of a sodium discharge lamp (Pfaender *et al.*, 2020). However, the tool is most often used to analyze video experiments concerning kinematics (Brown and Cox, 2009; Wee *et al.*, 2012; Gröber *et al.*, 2014; Wee *et al.*, 2015; and Onorato *et al.*, 2021). For example, Klein *et al.* (2014) showed how students recorded the motion of a ball thrown vertically from a moving skateboard on their smartphones, uploaded it into *Tracker*, and then used the visualization tool to track the trajectory of the ball (Fig. 21.2).

21.3.3 Augmented reality and mixed reality

Augmented reality (AR) and mixed reality (MR) comprise a class of visualization tools that generally aim for a real-time, immersive experience (Dede, 1995; Whitelock *et al.*, 1996; and Dede *et al.*, 1999). While related virtual reality (VR) tools are based completely on virtual, computer-based information (more consistent with Function II), AR and MR tools tend to overlay virtual images and information

---

[3] A platform that also hosts *Easy Java Simulations* (Esquembre, 2003; and Christian and Esquembre, 2007) and other visualization tools (https://www.compadre.org/osp/index.cfm).

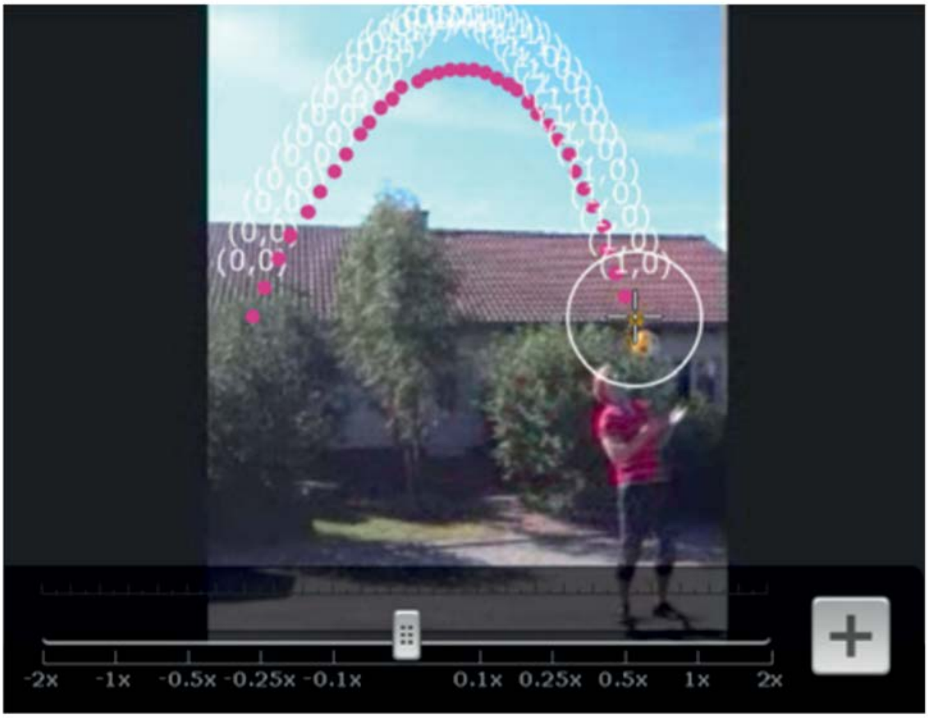


FIG. 21.2
Point-tracking the trajectory of a ball thrown vertically from a moving skateboard in *Tracker*. Video analysis of projectile motion using tablet computers as experimental tools. © IOP Publishing. Reproduced with permission. All rights reserved. (Taken from Klein *et al.*, Phys. Educ. **49**(1), 37–40 (2014). Copyright 2014 IOP Publishing).

onto physical objects, either through the view of a screen or directly projected on the objects themselves (consistent with Function I). Many AR solutions have been introduced in teaching, not least in science subjects (Cheng and Tsai, 2013). In a literature review of the advantages and challenges of AR in education, Akçayır and Akçayır (2017) identified many studies that reported positive effects on learning and students' attitudes. In line with Function I, the juxtaposition of real objects and virtual information in AR systems has been argued to reduce the cognitive load of integration of the different data inputs, though some AR/MR researchers caution about cognitive overload if students have to coordinate too much data/input at once (Cheng and Tsai, 2013).

In research on the MR environment *METeor*, researchers asked lower secondary school students to predict the path of an asteroid projected onto the floor as it was launched into motion (Lindgren *et al.*, 2016). In the context of mechanics, Enyedy *et al.* (2012) examined how students used an AR tool to playfully enact how a virtual object would respond to forces such as friction and pushes perpendicular to the motion. The students incorporated various symbols to represent the forces, some of which were more mathematically formal (e.g., arrows showing a push in a certain direction) and some of which were informal (e.g., an image of a surface with high friction). Such a mixture of symbols could be seen as exemplifying the mathematization made possible by visualization tools—i.e., blending of formal and inform domains that aligns with semi-formal modeling (Euler and Gregorcic, 2018).

21.3.3.1 Augmented reality example: Visualizing vector fields with AR

To externally visualize magnetic fields, Cai *et al.* (2017) introduced AR visualization tools in the teaching of magnetism in lower secondary schools. By integrating a model of the magnetic field into a Kinect® input device, the mathematized representations of magnetic field lines can be rendered around magnets that students move in space. Several other solutions for visualization of magnetic fields using augmented reality are available for teaching and learning purposes. For example, *MAGNA AR* (https://www.magna-ar.net/), a mobile-based AR tool, displays magnetic field vectors in real space and allows students to see and interact with 3D fields. *MAGNA AR* can be used to explore the static magnetic fields around a bar magnet (Fig. 21.3, left) or around a current-carrying loop of wire (Fig. 21.3, right).

Another AR visualization tool is presented by Yoon *et al.* (2018), where visitors in a science museum interact with a device called *Magnetic Maps*. Students who visited the museum played with real bar magnets while the interaction was captured by an overhead camera, digitized, and simultaneously fed back in real time. Magnetic force field lines appeared around the magnets on the computer screen, updating dynamically as the students moved the physical magnets.

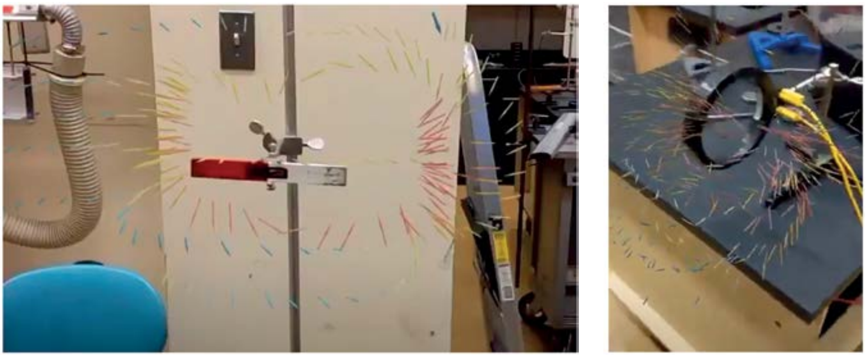


FIG. 21.3
Screenshots of the MAGNA AR visualization tool, superimposing the magnetic field vectors around a bar magnet (left) and around a current-carrying loop (right). Retrieved from https://www.magna-ar.net

21.4 FUNCTION II: BRIDGING IDEALIZED MODELS AND FORMAL REPRESENTATIONS

Having provided some examples of visualization tools that support mathematization through Function I, we now review some of the types of visualization tools in physics education that exemplify mathematization through Function II in some capacity: simulations, programming, microworlds, and educational games. As in the previous section, we briefly discuss each type of visualization tool and then provide an example of a visualization tool within that type used in the teaching and learning of physics. Consistent with Function II, the tools featured in this section dynamically depict formal representations as superimposed or alongside idealized models of physical phenomena.

21.4.1 Simulations

Simulations are pre-programmed models of phenomena that are depicted through an interactive pictorial/graphical interface. Especially compared to visualization tools such as programming environments (Sec. IV B) and microworlds (Sec. IV C), simulations are characteristically *constrained* in their design (Euler *et al.*, 2020), meaning they center on a specific phenomenon or mathematical concept and only present the aspects of that phenomenon or concept that is most relevant for students. This constrained nature of simulations has had meaningful consequences for the nature of the visualization/mathematization work done with them. For one, simulations tend to be bespoke visual renderings for each phenomenon that only includes visual features and controls that will directly contribute to the teaching and learning of the phenomenon in concern. Thus, the line of reasoning between the idealized models at hand and the mathematized representations associated with those models can remain relatively "uncluttered" for students. Scholars have found that students using simulations somewhat reliably discern the relevant connection between the specifically modeled phenomena and the formalism that describe them (Podolefsky *et al.*, 2010).

Simulations often allow manipulation of the scale of physics phenomena in both time and space. This is particularly useful in the content areas such as astronomy and quantum mechanics, both of which concern objects and processes at incomprehensibly large and small scales. Beyond the "resizing" of temporal scales, the manipulation of time scales can also be afforded by visualization tools through playback controls such as play, pause, and rewind.

21.4.1.1 Simulation example: PhET

PhET simulations (Perkins *et al.*, 2006) are among the most prominent examples of simulation tools in physics education. The suite of PhET simulations is adaptable to many education levels and has been implemented alongside many instructional approaches, ranging from lecture-based demonstrations (Correia *et al.*, 2019) to hands-on student activities/labs (Kohl and Finkelstein, 2005; and Moore

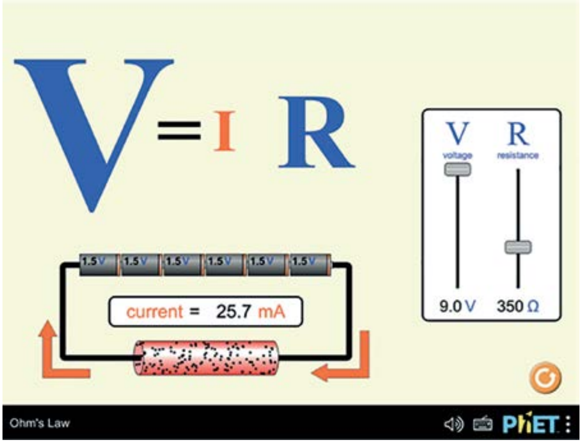


FIG. 21.4
The PhET visualization tool *Ohm's Law*.

et al., 2014). One innovation of the PhET project as compared to previous simulation-type visualization tools was in testing a set of design principles through student interviews rather than relying on the "designer's preferences or ease of coding" (Adams *et al.*, 2008, p. 552)—an approach which was later repeated with lower-secondary students to improve sim design for younger students (Paul *et al.*, 2013).

As with other simulation-type visualization tools, each PhET sim presents an interactive model of a single physics phenomenon or set of phenomena with interface controls for the disciplinary-relevant variables for said phenomenon/phenomena. For example, the *Ohm's Law* sim (Fig. 21.4) provides students with an abstract environment to visually experience the algebraic relationship between voltage, current, and resistance in the $V = IR$ form of Ohm's law and to connect those variables with physical features of circuits. Users interact with the sim by manipulating sliders for either voltage or resistance. As they do so, the corresponding variables V and R in the algebraic expression are scaled in visual size, and a cartoon of the physical circuit elements is augmented (with more or fewer batteries and/or more or fewer black dots in a cylindrical resistor). Synchronized with the manipulation of the voltage and resistance sliders, the sim dynamically updates the current in the system by scaling the visual size of the red variable I in the algebraic expression, scaling red arrows along the wires of the circuit diagram, and updating a decimal readout of the amperage labeled "current."

21.4.2 User-created simulations (programming)

Enabling students to create their own simulations through programming has been an aspirational goal for physics educators since the 1960s. Alfred Bork was an early advocate for developing courses designed to give students experience in solving physics problems by programming with FORTRAN

(Bork, 1963, 1967, 1968). In the late 1960s, Feurzeig and colleagues at MIT developed the LOGO computing language as an easier avenue into programming (Feurzeig *et al.*, 1969; and Papert, 1980).[4] Nonetheless, it was acknowledged by many physics education scholars as early as the 1970s that the student programming work with lower-level languages such as FORTRAN could entail little visualization beyond plotting functions (Bork and Ballard, 1972). Indeed, for decades it remained too great a task for the average student to generate their own visualizations.

Nonetheless, several programming-based physics education projects have revisited these visualization tools, such as M.U.P.P.E.T (MacDonald *et al.*, 1988; and Redish and Wilson, 1993) and CUPLE (Redish *et al.*, 1992; and Wilson, 1994). These newer projects took advantage of innovative languages such as PASCAL, C, and BASIC, though they ultimately lost momentum as viable curricular reforms. It was not until the creation of VPYTHON at Carnegie Melon in 2000 (Scherer *et al.*, 2000; and Sherwood, 2017) that programming in physics finally found an attainable route for most students to generate their own visualizations/simulations. Chabay and Sherwood soon thereafter finalized the *Matter & Interactions* curriculum (Chabay and Sherwood, 2002), which incorporated students' use of VPYTHON throughout (Chabay and Sherwood, 1999, 2004, 2008; Kohlmyer *et al.*, 2009; and Ding *et al.*, 2013). With entry points into programming becoming more readily available and with computers now ubiquitous in many societies, projects highlighting "computational" physics and programming are becoming increasingly common in physics education (e.g., Odden and Caballero, 2020).

21.4.2.1 User-created simulation example: GeoGebra

Though not a programming platform in the strictest sense, *GeoGebra* is a prime example of a visualization tool designed to empower students (and teachers) to create their own simulations (Hohenwarter and Fuchs, 2004). The main functionality of *GeoGebra* consists of an "algebra window" where mathematical expressions/equations can be entered by the user and a "graphical window" where the expressions/equations are dynamically rendered as a simulation. The key advantage of *GeoGebra*, especially when compared to the coding typically required to create a simulation, is its user-friendly design (Malgieri, Onorato, and De Ambrosis, 2014), a design that only requires an understanding of mathematics itself rather than specific programming expertise (Walsh, 2017). In the context of physics education, *GeoGebra* has been used by teachers from primary school to university to create simulations, augment real experiments, and/or directly involve students in the mathematizing physical phenomena (Solvang and Haglund, 2021).

Marciuc *et al.* (2016) had students use *GeoGebra* to create their own models of sliding motion on an inclined plane (Fig. 21.5). The mathematics was rendered dynamically in real-time such that the computer offered the students feedback throughout the activity and enabled a "constant confrontation between the meaning assigned to mathematical relations and the obtained images" (Marciuc *et al.*, 2016, p. 221).

---

[4] In fact, the LOGO language and the research/development that it inspired alongside Papert's (1980) book *Mindstorms* are historically more associated with microworlds. Still, it is relevant to note that this early microworld work was done through student programming.

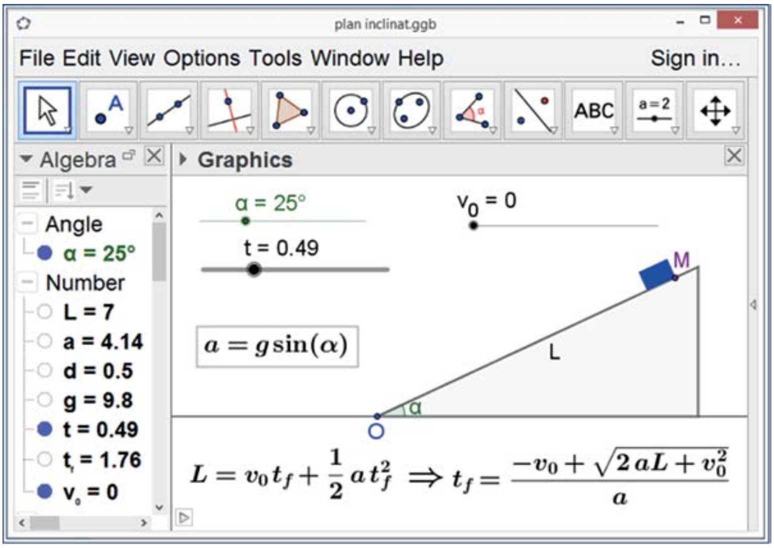


FIG. 21.5
GeoGebra screenshot showing a model of an inclined plane. Reproduced from Marcuic *et al.* (2016). "Learning physics by building computer models—movements on inclined planes," in *Proceedings of the 11th International Scientific Conference on ELearning and Software for Education* (Carol I National Defence University Publishing House, 2016), with the permission of the ADL ROMANIA. Copyright 2016 ADL ROMANIA.

In this way, *GeoGebra* can be seen as a visualization tool that encourages students to interrogate their interpretations of mathematical formulas in the context of physics—moving *from* mathematics *to* the simulated model—a key to the back-and-forth of mathematization of Function II. One key advantage highlighted by researchers is that *GeoGebra* "makes the mathematical models behind the simulations completely transparent and easily accessible to the user, and avoids producing the impression that complex and exotic algorithms are at work" (Malgieri, Onorato, and De Ambrosis, 2014, p. 19).

21.4.3 Microworlds

Microworlds are visualization tools that offer more opportunities for creative engagement than what is typical of simulation-type visualization tools. In contrast to the intentionally-constrained nature of simulations, where students tend to explore a single phenomenon via a model someone else has built, microworlds are "less-constrained" (Euler *et al.*, 2020) software that allow students to build their own models of a diverse range of physical phenomena within the same software (Laurillard, 2002). Seymour Papert (1980) used microworld to describe a family of programming languages called LOGO systems (Feurzeig *et al.*, 1969), which were intended to foster creativity by immersing learners in an environment that provides them with "building blocks" of mathematical language and reasoning.

Microworlds also became the posterchild for Papert's theory of learning called *constructionism*, "learning-by-making" philosophy that posits that learning happens "especially felicitously in a context where the learner is consciously engaged in constructing a public entity" (Papert, 1991, p. 1). By the middle of the 1980s, many other microworld-type visualization tools soon followed LOGO, such as STELLA (Richmond, 1985, 1992) and *Boxer* (diSessa, 1986; diSessa and Abelson, 1986; Adams and diSessa, 1991; and diSessa, Abelson *et al.*, 1991), and *ThinkerTools* (White and Horwitz, 1987; and White, 1992, 1993). More modern microworld-type visualization include *Algodoo* (Algoryx Simulation AB, 2011) and *Fizika* (Radnai *et al.*, 2019). These visualization tools generally incorporate formal, mathematical representations throughout their less-constrained design, making them particularly useful as tools for encouraging students to construct their own models and creatively mathematize in the context of physics.

21.4.3.1 Microworld example: Algodoo

Algodoo is a 2D Newtonian microworld wherein users can build their own virtual experiments, scenarios, and machines using simple geometric shapes and functional parts such as springs, chains, axels, and hinges (Gregorcic and Bodin, 2017; and Euler and Gregorcic, 2019). In contrast to the models made in programming tools like VPYTHON, where the necessary mathematics is chosen and imported by the user, models in *Algodoo* can be created without explicit use of formal mathematics since *Algodoo* expects the user to draw the components of the model directly into the 2D graphical environment. Physical constants such as gravitational acceleration and air resistance can be manipulated via sliders, allowing for a variety of uses across many physical phenomena. Additional visual tools allow the augmenting and supplementing of virtual objects with formal, mathematical representations (e.g., vector arrows, plots; see Fig. 21.6). As with many of the visualization tools discussed in this chapter, the dynamic nature of these representations allows users to observe their evolution synchronized to the evolution of the created models (in real time, slowed down/sped up, or intermittently paused/replayed). Beyond the built-in mathematical representations, students' engagement with formal physics concepts can be encouraged via the manipulation of model parameters corresponding to the physical properties of objects and the environment (Euler *et al.*, 2020).

Research that features *Algodoo* is diverse, addressing, for example, students' epistemic beliefs in a computer-simulated problem solving (Lindfors *et al.*, 2019), the combination of *Algodoo* with an interactive whiteboard in high-school physics classrooms (Gregorcic, Etkina *et al.*, 2017), as well as more theoretical publications on the topics of variation theory of learning (Euler *et al.*, 2020), multimodal communication (Gregorcic, Planinsic *et al.*, 2017), conceptual blending (Gregorcic and Haglund, 2021), and the interplay of physics and mathematics in students' learning activities (Euler and Gregorcic, 2019).

Though the mathematical formalism of Newtonian mechanics are embedded in the code of *Algodoo* and users have access to many relevant physics representations, the underlying computational model of physics is kept hidden from the user. Herein lies the tradeoff inherent to most microworld-like visualization tools: the technology is intuitive to use, yet this ease-of-use comes at the price of giving

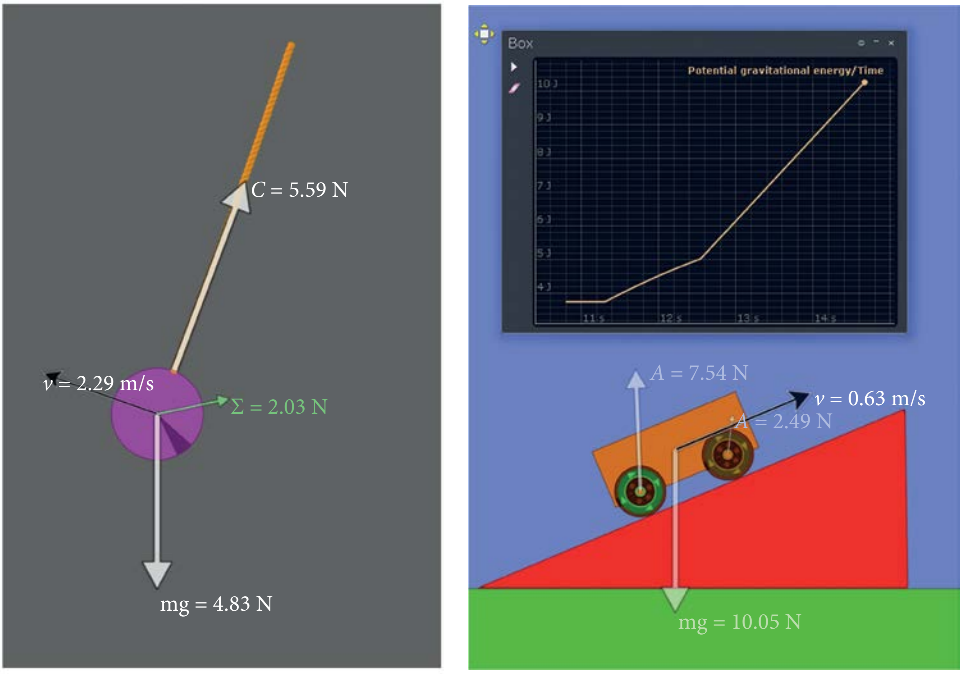
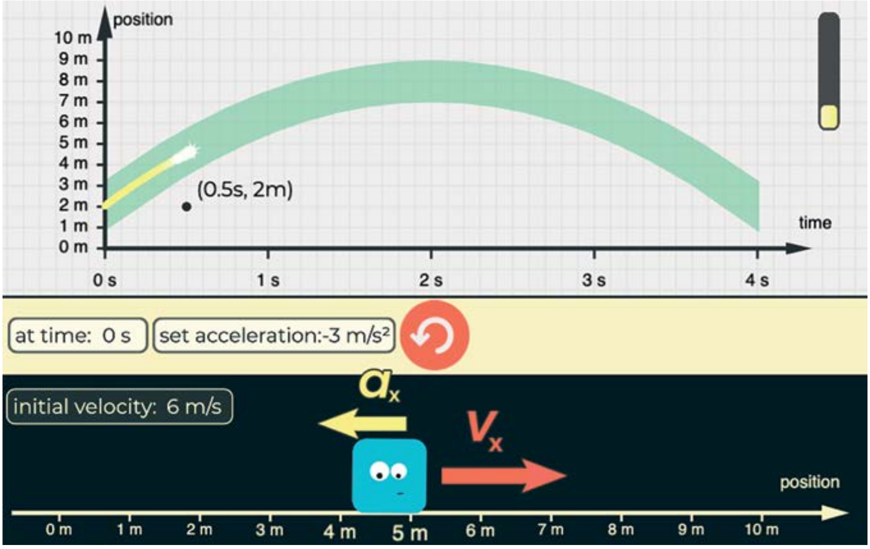


FIG. 21.6
A modeled pendulum in *Algodoo* with dynamic vector arrows as they change throughout the pendulum swing—velocity in black, external forces on the bob in white, and the sum of all external forces in green (left). A constructed "car" in *Algodoo* with motors on the front and rear wheels ascends a slope (right). Dynamic vector arrows are displayed for the orange box, while the graph above shows the time dependence of the box's gravitational potential energy.

limited insight into many of the details of the modeling process (Gregorcic and Haglund, 2021). Other visualization tools like *GeoGebra* may show more of the underlying mathematical formalism but often do so without as intuitive a connection to physical objects and phenomena.

21.4.4 Educational games

In recent decades, gamification, the "process of introducing game mechanics, dynamics, and frameworks to promote desired behaviors" (Lee and Hammer, 2011, p. 1), has become an ever more prevalent topic in education. Effective games motivate players, in part by allowing them to practice required skills repeatedly and thus help them frame failure as a productive part of learning. The process by which games influence learning is often linked to the idea of internal visualization, where students learn to "simulate" the game dynamics in their mind's eye (Glenberg, 1997). Thus, if a game is well aligned in its content with given physics curriculum learning goals (such as the use of formal representations), there is potential to leverage playing a game to learn physics.



at time: 0 s

set acceleration:-3 m/s²

initial velocity: 6 m/s

FIG. 21.7
A screenshot of *Motion Mapper*.

Educational games comprise a diverse set of visualization tools distinct from other digital environments such as simulations and microworlds in that they are explicitly goal-directed. In some cases, the introduction of goals can change the nature of a tool from a simulation to a game (e.g., see the gamified version of *Graphs & Tracks*[5]). Games can be further categorized between those made for educational purposes—e.g., *A Slower Speed of Light* (Kortemeyer, 2019) and games available at theuniverseandmore.com—and those not made explicitly for educational purposes but nevertheless being used in education—e.g., *Angry Birds*, *Kerbal Space Programme*, *Universe Sandbox*. Alternatively, and especially relevant for the purpose of this section, we can stratify the games according to their degree of inclusion of formal representations. Most games, especially those not designed for educational use, do not include formal representations. On the other extreme, games such as *Motion Mapper* are built explicitly with the purpose of teaching the use and understanding of formal representations in physics.

21.4.4.1 Educational game example: Motion Mapper

Motion Mapper is a game with an explicit aim to strengthen students' ability to interpret kinematic graphs by linking them to perceivable motion in one dimension (see Fig. 21.7). The game has two modes: one where the user must figure out and enter numerical values of kinematic quantities (initial position, velocity, acceleration) to match a predetermined graph, and another where the user needs to use their

---

[5] http://graphsandtracks.com/

mouse or touch screen to move a character in accordance with predetermined kinematics graphs. In this way, the formal representations are linked to both the visual perception of motion in space along one dimension, as well as to the kinesthetic perception of the motion of one's hand via proprioception (perception of the position of one's own body). In this way, the game also makes use of the mechanisms of embodiment (Gee, 2008; Euler *et al.*, 2019; and Gregorcic and Haglund, 2021) for enhancing learning.

21.5 SUMMARIZING THE HISTORY OF VISUALIZATION TOOLS IN PHYSICS EDUCATION

Having reviewed several ways that mathematization in physics can be functionally supported through visualization tools, we now historically summarize the development of visualization tools in PER. Our synopsis of this history appears in Fig. 21.8. The horizontal bands of the timeline are organized to show

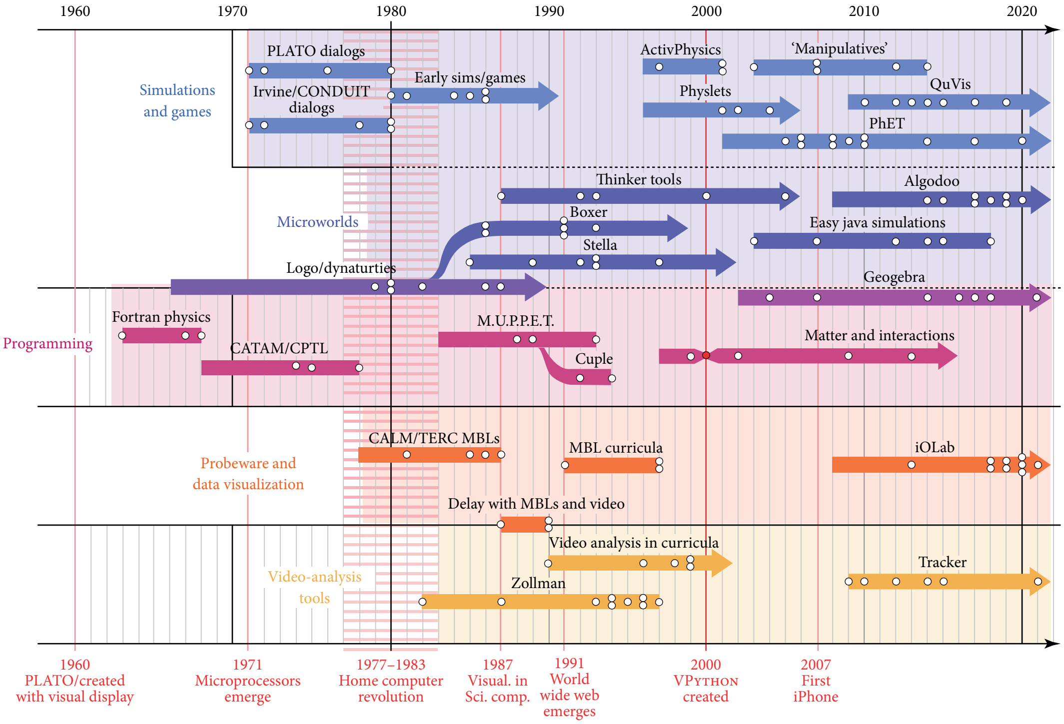


FIG. 21.8
Summary figure for the developmental history of visualization tools in PER. Key clusters of work have been collected and labeled (see Appendix).

some of the major strands of work within simulations & games, microworlds, programming, probeware & data visualization, and video-analysis tools. Dots along each strand represent publications (detailed in the Appendix) selected to show a spread of the work done around each strand through time. Some of the histories and interrelations of the strands of work involving these visualization tools have been included in the body of this chapter, but a full account of the topic is beyond the scope of this section (see Euler, in preparation). Still, it is important to emphasize that much of the work involving visualization tools today has descended from seeds of innovation cultivated as far back as the 1960s. For those wanting to see more specifics about each tool, see the resource letter from Euler *et al.* (in preparation).

21.6 CONCLUSIONS

We examined a collection of commonly used visualization tools in PER and discussed how they support mathematization. Specifically, we have reviewed how visualization tools can act as semi-formal bridges between physical phenomena, idealized phenomena, and mathematical formalism by making formal ideas more readily relatable to physical experiences. Our synthesis has collected a wide variety of efforts across PER history, simultaneously showing some of the growing efforts among PER scholars to utilize digital technologies for visualization but also demonstrating the overall lack of coherence for this topic of visualization and mathematization despite its fundamentality to physics teaching and learning. Future work in PER would do well to explicitly contend with how visualization tools like those reviewed above can facilitate mathematization and, conversely, how efforts to support students' movement back-and-forth between the physical world and formal mathematics might be strategically scaffolded by digital technologies.

APPENDIX—LEGEND FOR REFERENCES IN FIG. 21.8

Simulations and games

PLATO dialogs: Bennet (1972); Kane and Sherwood (1980); Sherwood (1971); and Smith and Sherwood (1976). **Irvine/CONDUIT dialogs:** Bork (1978, 1980); Bork and Robson (1972); Bork and Sherman (1971); and Peters (1980). **Early sims and games:** Hewson (1985); Reed and Saavedra (1986); Trowbridge and McDermott (1980, 1981); Zietsman and Hewson (1986); and Zollman (1984). ***ActivPhysics*:** Furtak and Ohno (2001); Van Heuvelen (2001, 1997). ***Physlets*:** Belloni *et al.* (2004); Christian and Belloni (2001); and Dancy *et al.* (2002). **"Manipulatives":** Chini *et al.* (2012); Klahr *et al.* (2007); Triona and Klahr (2003); Zacharia (2007); and Zacharia and de Jong (2014). **PhET:** Adams *et al.* (2006); Finkelstein *et al.* (2005), (2006); López-Tavares *et al.* (2020); McKagan *et al.* (2008); Moore *et al.* (2014); Perkins and Moore (2017); Podolefsky *et al.* (2010, 2009); Wieman *et al.* (2010, 2008). ***QuVis*:** Kohnle (2014, 2013); Kohnle *et al.* (2010, 2012, 2015); Kohnle and Rizzoli (2017); and Passante and Kohnle (2019).

Microworlds

Logo/Dynaturtles: Brna (1987); Clements (1986); diSessa (1980, 1982); Papert *et al.* (1979); and Papert (1980). **STELLA:** Niedderer *et al.* (1989, 1997); Richmond (1992, 1985); Schecker (1993); and Tinker (1993). ***Boxer*:** Adams and diSessa (1991); diSessa (1986); diSessa, Abelson *et al.* (1991); diSessa, Hammer *et al.* (1991); diSessa and Abelson (1986); and Sherin *et al.* (1993). ***ThinkerTools*:** Schwarz and White (2005); White (1992, 1993); White and Frederiksen (2000); and White and Horwitz (1987). ***Easy Java Simulations*:** Christian and Esquembre (2007); Esquembre (2003); Esquembre *et al.* (2018); Wee (2012); Wee *et al.* (2015a); and Wee and Ning (2014). ***Algodoo*:** da Silva *et al.* (2014); Euler *et al.* (2020); Euler and Gregorcic (2019); Gregorcic (2015b); Gregorcic, Planinsic *et al.* (2017); Gregorcic and Bodin (2017); Radnai *et al.* (2019); and Vliora *et al.* (2018).

Programming

Fortran: Bork (1963, 1967, 1968). **CATAM/CPTL:** Harding (1974, 1975); Hinton (1978). **M.U.P.P.E.T.:** MacDonald *et al.* (1988); Redish and Wilson (1993); and Wilson and Redish (1989). **CUPLE:** Redish *et al.* (1992); and Wilson (1994). ***Matter and Interactions*:** Chabay and Sherwood (1999, 2002); Ding *et al.* (2013); Kohlmyer *et al.* (2009); and Scherer *et al.* (2000). ***GeoGebra*:** Hohenwarter and Fuchs (2004); Hohenwarter and Jones (2007); Malgieri *et al.* (2014); Marciuc *et al.* (2016); Solvang and Haglund (2021, 2018); and Walsh (2017).

Probeware and data visualization

CALM/TERC MBLs: Barclay (1986); Mokros (1985); Thornton (1987); and Tinker (1981). **Delay with MBLs and video:** Beichner (1990); Brasell (1987); and Thornton and Sokoloff (1990). **MBL curricula:** Laws (1991); Sokoloff and Thornton (1997); and Thornton and Sokoloff (1997). **iOLab:** Ansell (2020); Ansell and Selen (2018); Bodegom *et al.* (2019); Leblond and Hicks (2021); Nair and Sawtelle (2018); Selen (2013); Volkwyn *et al.* (2019); Volkwyn *et al.* (2020a); and Volkwyn *et al.* (2020b).

Video-analysis tools

Zollman: Brungardt and Zollman (1995); Chaudhury *et al.* (1994); Dengler *et al.* (1993); Escalada *et al.* (1996); Escalada and Zollman (1997); Fuller *et al.* (1982); Zollman *et al.* (1987); Zollman (1996); and Zollman and Fuller (1994). **Video-analysis in curricula:** Beichner (1996); Beichner *et al.* (1999); Beichner and Abbott (1999); Cadmus (1990); Laws and Pfister (1998). ***Tracker*:** Brown and Cox (2009); Chanpichai *et al.* (2010); Malgieri, Onorato, Mascheretti *et al.* (2014); Onorato *et al.* (2021); Wee *et al.* (2012); and Wee *et al.* (2015b).

REFERENCES

Adams, S. T. and diSessa, A. A., J. Math. Behav. **10**(1), 79–89 (1991).
Adams, W. K. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **2**(1), 010101 (2006).
Adams, W. K. *et al.*, J. Interac. Learn. Res. **19**(4), 551–577 (2008).
Ainsworth, S., Learn. Instr. **16**(3), 183–198 (2006).
Ainsworth, S., Comput. Educ. **33**(2–3), 131–152 (1999).
Airey, J. and Linder, C., *Multiple Representations in Physics Education*, edited by D. F. Treagust *et al.* (Springer International Publishing, 2017), pp. 95–122.
Akçayır, M. and Akçayır, G., Educ. Res. Rev. **20**, 1–11 (2017).
Algoryx Simulation AB, see www.algodoo.com for "Algodoo" (2011),
Alibali, M. W. and Nathan, M. J., J. Learn. Sci. **21**(2), 247–286 (2012).
Ansell, K., doctoral dissertation (University of Illinois at Urbana-Champaign, 2020), see http://hdl.handle.net/2142/107872
Ansell, K. and Selen, M., 2017 Physics Education Research Conference Proceedings (American Association of Physics Teachers, 2018), pp. 40–43.
Barclay, W. L., 7th National Educational Computing Conference (Technical Education Research Center, 1986), pp. 1–10.
Barsalou, L. W., Annu. Rev. Psychol. **59**(1), 617–645 (2008).
Beichner, R. J. and Abbott, D. S., *Video-Based Labs for Introductory Physics. November*, 101–104 (1999).
Beichner, R. J., J. Res. Sci. Teach. **27**(8), 803–815 (1990).
Beichner, R. J., Am. J. Phys. **64**(10), 1272–1277 (1996).
Beichner, R. J. *et al.*, Am. J. Phys. **67**(7), S16–S24 (1999).
Belloni, M. *et al.*, Phys. Teach. **42**(5), 284–290 (2004).
Bennet, C. D., Conference on Computers in Undergraduate Curricula (The Commission on College Physics, 1972).
Bing, T. J. and Redish, E. F., Phys. Rev. Spec. Top. Phys. Educ. Res. **5**(2), 020108 (2009).
Black, K. E. *et al.*, AIP Conf. Proc. **951**, 53–56 (2007).
Blum, W. and Borromeo, R., J. Math. Model. Appl. **1**(1), 45–58 (2009).
Bodegom, E. *et al.*, Phys. Teach. **57**(6), 382–386 (2019).
Bork, A. M. and Ballard, R., see https://files.eric.ed.gov/fulltext/ED060616.pdf for "Computer graphics and physics teaching" (1972).
Bork, A. M. and Robson, J., Am. J. Phys. **40**, 1288–1294 (1972).
Bork, A. M. and Sherman, N., Am. J. Phys. **39**(2), 137–143 (1971).
Bork, A. M., Am. J. Phys. **31**, 364–368 (1963).
Bork, A. M., Am. J. Phys. **46**(8), 796–800 (1978).
Bork, A. M., Am. J. Phys. **36**(10), 907 (1968).
Bork, A. M., Comput. Educ. **4**(1), 37–57 (1980).
Bork, A. M., *Fortran for Physics* (Addison-Wesley Publishing Co, 1967).
Brahmia, S. M., *Mathematization in Introductory Physics* (Rutgers University, 2014).
Brasell, H., J. Res. Sci. Teach. **24**(4), 385–395 (1987).
Brna, P., Instr. Sci. **16**(4), 351–379 (1987).
Brown, D. and Cox, A. J., Phys. Teach. **47**(3), 145–150 (2009).
Brown, D., *Paper Presented at the 2008 AAPT Summer Meeting*, Edmonton (American Association of Physics Teachers, 2008).
Brungardt, J. B. and Zollman, D. A., J. Res. Sci. Teach. **32**(8), 855–869 (1995).
Bryan, J. A., Learn. Lead. Technol. **32**(6), 22–24 (2005).
Cadmus, R. R., Am. J. Phys. **58**(4), 397–399 (1990).
Cai, S. *et al.*, Interact. Learn. Environ. **25**(6), 778–791 (2017).
Card, S. K. *et al.*, *The Psychology of Human-Computer Interaction* (Taylor & Francis, 1983).
Chabay, R. W. and Sherwood, B. A., Am. J. Phys. **67**(12), 1045–1050 (1999).
Chabay, R. W. and Sherwood, B. A., Am. J. Phys. **76**(4), 307–313 (2008).
Chabay, R. W. and Sherwood, B. A., Am. J. Phys. **72**(4), 439–445 (2004).
Chabay, R. W. and Sherwood, B. A., *Matter & Interactions I. Modern Mechanics* (Wiley, 2002).
Chandler, P. and Sweller, J., Cogn. Instr. **8**(4), 293–332 (1991).
Chandler, P. and Sweller, J., Br. J. Educ. Psychol. **62**(2), 233–246 (1992).
Chanpichai, N. *et al.*, AIP Conf. Proc. **1263**, 212–215 (2010).
Chaudhury, S. R. *et al.*, Comput. Phys. **8**(5), 518 (1994).
Chen, Z. and Gladding, G., Phys. Rev. Spec. Top. Phys. Educ. Res. **10**(1), 1–24 (2014).
Cheng, K.-H. and Tsai, C.-C., J. Sci. Educ. Technol. **22**(4), 449–462 (2013).
Chini, J. J. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **8**(1), 010113 (2012).
Christian, W. and Belloni, M., *Physlets: Teaching Physics with Interactive Curricular Material* (Prentice Hall, Inc., 2001).

Christian, W. and Esquembre, F., Phys. Teac. **45**(8), 475–480 (2007).

Clement, J. J., *Creative Model Construction in Scientists and Students: The Role of Imagery, Analogy, and Mental Simulation*, edited by J. J. Clement (Springer, Netherlands, 2008).

Clements, D. H., J. Educ. Psychol. **78**(4), 309–318 (1986).

Collins, A. and Ferguson, W., Educ. Psychol. **28**(1), 25–42 (1993).

Correia, A.-P. *et al.*, Res. Sci. Technol. Educ. **37**(2), 193–217 (2019).

da Silva, S. L. *et al.*, Exatas Online **5**(2), 28–39 (2014), see http://arxiv.org/abs/1409.1621

Dancy, M. H. *et al.*, Phys. Teach. **40**(8), 494–499 (2002).

Dede, C., Educ. Technol. **35**(5), 46–52 (1995).

Dede, C. *et al.*, *Modeling and Simulation in Science and Mathematics Education* (Springer, New York, 1999), pp. 282–319.

Defanti, T. A. and Brown, M. D., *Advances in Computers*, edited by B. H. McCormick *et al.* (Academic Press, 1991), Vol. 33, Issue C, pp. 247–307.

Dengler, R. *et al.*, Comput. Phys. **7**(4), 393 (1993).

Ding, L. *et al.*, J. Res. Sci. Teach. **50**(6), 722–747 (2013).

diSessa, A. A. and Abelson, H., Commun. ACM **29**(9), 859–868 (1986).

diSessa, A. A., Comput. Educ. **4**(1), 67–75 (1980).

diSessa, A. A., Inst Sci **14**, 207–227.

diSessa, A. A., *Constructivism in the Computer Age*, edited by G. E. Forman and P. B. Pufall (Lawrence Erlbaum Publishers, 1988), pp. 49–70.

diSessa, A. A., Cogn. Sci. **6**(1), 37–75 (1982).

diSessa, A. A. *et al.*, J. Math. Behav. **10**(1), 3–15 (1991a).

diSessa, A. A. *et al.*, J. Math. Behav. **10**(2), 117–160 (1991b).

Dreyfus, B. W. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 020141 (2017).

Enyedy, N. *et al.*, Int. J. Comput. Support. Collab. Learn. **7**(3), 347–378 (2012).

Eriksson, U., *Reading the Sky. From Starspots to Spotting Stars* (Uppsala University, 2014).

Escalada, L. T. and Zollman, D. A., J. Res. Sci. Teach. **34**(5), 467–489 (1997).

Escalada, L. T. *et al.*, J. Educ. Multimedia Hypermedia **5**(1), 73–97 (1996).

Esquembre, F., *The IEEE Region 8 EUROCON 2003: Computer as a Tool* (IEEE, 2003), Vol. 1, pp. 20–23.

Esquembre, F. *et al.*, Am. J. Phys. **86**(1), 54–67 (2018).

Euler, E., *Educational technology in physics I: A review in historical periods* (unpublished).

Euler, E. and Gregorcic, B., *Mathematics in Physics Education*, edited by G. Pospiech *et al.*, (Springer International Publishing, 2019), pp. 355–385.

Euler, E. and Gregorcic, B., *2017 Physics Education Research Conference Proceedings* (American Association of Physics Teachers, 2018), pp. 128–131.

Euler, E. *et al.*, Eur. J. Phys. **41**(4), 045705 (2020).

Euler, E. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(1), 010134 (2019).

Euler, E. *et al.*, *Educational technology in physics II: Resource Letter* (unpublished).

Feurzeig, W. *et al.* (1969). *Programming Languages as a Conceptual Framework for Teaching Mathematics. Final Report on the First Fifteen Months of the LOGO Project.*

Finkelstein, N. D. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **1**(1), 1–8 (2005).

Finkelstein, N. D. *et al.*, MERLOT J. Online Learn. Teach. **2**(3), 110–121 (2006).

Fredlund, T. *et al.*, Eur. J. Phys. **36**(5), 055001 (2015).

Fuller, R. G. *et al.*, *The Puzzle of the Tacoma Narrows Bridge Collapse (Videodisc)* (University of Nebraska, 1982).

Furtak, T. E. and Ohno, T. R., Phys. Teach. **39**(9), 534–538 (2001).

Galton, F., *Inquiries Into Human Faculty and Its Development* (MacMillan Co, 1883), pp. 83–114.

Ganguly, I., *Eyes on the Future: Converging Images, Ideas and Instruction—Selected Readings From the Annual Conference on the International Visual Literacy Association* (ERIC Clearinghouse, 1995), pp. 241–250.

Gee, J. P., Games Cult. **3**(3–4), 253–263 (2008).

Gingras, Y., Hist. Sci. **39**(4), 383–416 (2001).

Glenberg, A. M., Behav. Brain Sci. **20**(1), 1–19 (1997).

Gregorcic, B. and Bodin, M., Phys. Teach. **55**, 25–28 (2017).

Gregorcic, B. and Haglund, J., Res. Sci. Educ. **51**(2), 235–275 (2021).

Gregorcic, B., Phys. Educ. **50**(5), 511–515 (2015b).

Gregorcic, B. *et al.*, Res. Sci. Educ. **48**(2), 1–25 (2017a).

Gregorcic, B., *Investigating and Applying Advantages of Interactive Whiteboards in Physics Instruction* (University of Ljubljana, 2015a).

Gregorcic, B. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 1–17 (2017b).

Gröber, S. *et al.*, Eur. J. Phys. **35**(5), 055019 (2014).

Hammer, D. *et al.*, *Transfer of Learning From a Multidisciplinary Perspective*, edited by J. P. Mestre (Information Age Pub, 2005), pp. 89–119.

Harding, R. D., Int. J. Math. Educ. Sci. Technol. **5**(3–4), 447–455 (1974).

Harding, R. D., *Computer Assisted Learning in the United Kingdom*, edited by R. Hooper and I. Toye (Councils and Education Press, 1975).

Hestenes, D., Am. J. Phys. **60**(8), 732–748 (1992).

Hestenes, D., Am. J. Phys. **55**(5), 440–454 (1987).

Hewson, P. W., Am. J. Phys. **53**(7), 684–690 (1985).

Hinton, T., Comput. Educ. **2**, 71–88 (1978).

Hodge, R. and Kress, G., *Social Semiotics* (Polity Press, 1988).
Hohenwarter, M. and Fuchs, K., *Computer Algebra Systems and Dynamic Geometry Systems in Mathematics Teaching Conference* (Sapientia Hungarian University of Transylvania, 2004), pp. 1–6.
Hohenwarter, M. and Jones, K., Proc. Br. Soc. Res. Learn. Math. **27**(3), 126–131 (2007).
Hollan, J. *et al.*, ACM Trans. Comput. Hum. Interact. **7**(2), 174–196 (2000).
Kane, D. and Sherwood, B. A., Comput. Educ. **4**(1), 15–36 (1980).
Klahr, D. *et al.*, J. Res. Sci. Teach. **44**(1), 183–203 (2007).
Klein, P. *et al.*, Phys. Educ. **49**(1), 37–40 (2014).
Kohl, P. B. and Finkelstein, N. D., Phys. Rev. Spec. Top. Phys. Educ. Res. **1**(1), 010104 (2005).
Kohlmyer, M. A. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **5**(2), 020105 (2009).
Kohnle, A. and Rizzoli, A., Eur. J. Phys. **38**(3), 035403 (2017).
Kohnle, A., *Multimedia in Physics Teaching and Learning Conference (MPTL18)* (GIREP, 2013), pp. 7–12.
Kohnle, A., *Proceedings From the 2014 GIREP-MPTL Conference* edited by C, Fazio and R, M. Sperandeo-Mineo (GIREP, 2014), pp. 29–39.
Kohnle, A. *et al.*, in *2014 Physics Education Research Conference Proceedings*, (American Association of Physics Teachers, 2015), pp. 139–142.
Kohnle, A. *et al.*, Am. J. Phys. **80**(2), 148–153 (2012).
Kohnle, A. *et al.*, Eur. J. Phys. **31**(6), 1441–1455 (2010).
Kortemeyer, G., J. Phys. Conf. Ser. **1286**(1), 012048 (2019).
Kuhn, T. S., *The Structure of Scientific Revolutions (Second)* (University of Chicago Press, 1970).
Latour, B., *Representation in Scientific Practice*, edited by M. Lynch, and S. Woolgar (MIT Press, 1990), pp. 19–68.
Laurillard, D., *Rethinking University Teaching: A Conversational Framework for the Effective use of Learning Technologies*, 2nd Editio (Routledge, 2002).
Laws, P. and Pfister, H., Phys. Teach. **36**(5), 282–287 (1998).
Laws, P. W., Phys. Today **44**(12), 24–31 (1991).
Leblond, L. and Hicks, M., Phys. Teach. **59**(5), 351–355 (2021).
Lee, J. J. and Hammer, J., Acad. Exch. Q. **15**, 2 (2011).
Lee, K. M. *et al.*, J. Sci. Educ. Technol. **13**(1), 81–88 (2004).
Lindfors, M. *et al.*, Scand. J. Educ. Res. **63**(1), 124–144 (2019).
Lindgren, R. *et al.*, Comput. Educ. **95**, 174–187 (2016).
Lingefjärd, T. and Ghosh, J. B., Far East J. Math. Educ. **16**(3), 271–297 (2016).
López-Tavares, D. B. *et al.*, *2019 Physics Education Research Conference Proceedings* (American Association of Physics Teachers, 2020), pp. 579–584.
Lynch, M., Human Studies **11**(2/3), 201–234 (1988).
MacDonald, W. M. *et al.*, Comput. Phys. **2**(4), 23–30 (1988).
Malgieri, M. *et al.*, Eur. J. Phys. **35**(5), 055024 (2014a).
Malgieri, M. *et al.*, Phys. Educ. **49**(5), 500–511 (2014b).
Marciuc, D *et al.*, *Proceedings of the 11th International Scientific Conference on ELearning and Software for Education* (Carol I National Defence University Publishing House, 2016).
Mayer, R. E. and Moreno, R., Educ. Psychol. **38**(1), 43–52 (2003).
Mayer, R. E., *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer (Cambridge University Press, 2014), pp. 43–71.
Mayer, R. E., Educ. Psychol. **32**(1), 1–19 (1997).
McKagan, S. B. *et al.*, Am. J. Phys. **76**(4), 406–417 (2008).
Mešić, V. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(1), 010135 (2016).
Mokros, J. R., *Paper Submitted to the 1986 Annual Meeting of the National Association for Research in Science Teaching* (National Association for Research in Science Teaching, 1985).
Mokros, J. and Tinker, R., J. Res. Sci. Teach. **24**, 369–383 (1987).
Monaghan, J. M. and Clement, J. J., Int. J. Sci. Educ. **21**(9), 921–944 (1999).
Moore, E. B. *et al.*, J. Chem. Educ. **91**(8), 1191–1197 (2014).
Nair, A. and Sawtelle, V., Phys. Teach. **56**(8), 512–514 (2018).
Niedderer, H. *et al.* (1989). *Computers in physics education: First project report.*
Niedderer, H. *et al.*, AIP Conf. Proc. **399**(1997), 659–668 (1997).
Niss, M., Int. J. Sci. Math. Educ. **15**(8), 1441–1462 (2017).
Odden, T. O. B., and Caballero, M. D., *Physics Education Research Conference Proceedings* (American Association of Physics Teachers, 2020), pp. 429–434.
Onorato, P. *et al.*, Phys. Educ. **56**(4), 045007 (2021).
Papert, S *et al.*, *The Brookline LOGO Project. Final Report. Part II: Project Summary and Data Analysis* (1979).
Papert, S., *Constructionism*, edited by I. Harel and S. Papert (Ablex Publishing Corporation, 1991), pp. 1–11.
Papert, S., *Mindstorms: Children, Computers and Powerful Ideas* (Basic Books, 1980).
Passante, G. and Kohnle, A., Phys. Rev. Phys. Educ. Res. **15**(1), 010110 (2019).
Paul, A. *et al.*, AIP Conf. Proc. **1513**, 302–305 (2013).
Perkins, K. K. and Moore, E. B., *Physics Education Research Conference Proceedings (PERC)* (American Association of Physics Teachers, 2017), pp. 296–299.

Perkins, K. K. *et al.*, Phys. Teach. **44**(1), 18–23 (2006).
Peters, H. J., Comput. Educ. **4**(1), 1–9 (1980).
Pfaender, J. *et al.*, Phys. Educ. **55**(3), 033005 (2020).
Phillips, L. M. *et al.*, *Visualization in Mathematics, Reading and Science Education* (Springer, Netherlands, 2010), Vol. 5.
Podolefsky, N. S. *et al.*, AIP Conf. Proc. **1179**(November), 229–232 (2009).
Podolefsky, N. S. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **6**(2), 020117 (2010).
Radnai, T. *et al.*, J. Phys. Conf. Ser. **1223**(1), 012006 (2019).
Redish, E. F. and Bing, T. J., *Proceedings From the GIREP-EPEC & PHEC 2009 International Conference*, edited by D. Raine *et al.*, (GIREP, 2009), pp. 71–76, see http://physics.le.ac.uk/girep2009/ConferenceProceedings/GIREP2009_ConferenceProceedings_Volume2.pdf#page=85
Redish, E. F. and Wilson, J. M., Am. J. Phys. **61**(3), 222–232 (1993).
Redish, E. F., in *Conference, World View on Physics Education in 2005 Focusing on Change* (International Commission on Physics Education 2005), pp. 1–10.
Redish, E. F. *et al.*, Am. J. Phys. **66**(3), 212–224 (1998).
Redish, E. F. *et al.*, *Sociomedia: Multimedia, Hypermedia and the Social Construction of Knowledge*, edited by E. Barret (MIT Press, 1992), pp. 219–256.
Reed, S. K. and Saavedra, N. C., Cogn. Instr. **3**(1), 31–62 (1986).
Reiner, M., Int. J. Sci. Educ. **20**(9), 1043–1058 (1998).
Richmond, B. M., *Proceedings of the 1985 International System Dynamics Conference* (Systems Dynamic Society, 1985) pp. 706–718.
Richmond, B. M., *An Introduction to Systems Thinking: STELLA Software* (High Performance Systems, Inc., 1992).
Rieber, L. P., Educ. Technol. Res. Dev. **44**(2), 43–58 (1996).
Rosengrant, D. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **5**(1), 010108 (2009).
Roth, W.-M. and Lee, Y.-J., Rev. Educ. Res. **77**(2), 186–232 (2007).
Ryan, Q. X *et al.*, *Physics Education Research Conference Proceedings* (American Association of Physics Teachers, 2018), pp. 344–347.
Samuelsson, C. R. (2020). *Reasoning with thermal cameras settings in higher education*.
Schecker, H. P., Phys. Educ. **28**(2), 102–106 (1993).
Scherer, D. *et al.*, Comput. Sci. Eng. **2**(5), 56–62 (2000).
Schnotz, W. and Kürschner, C., Instr. Sci. **36**(3), 175–190 (2008).
Schnotz, W., *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer (Cambridge University Press, 2014), pp. 72–103.
Schwartz, J. L. and Taylor, E. F., *American Federation of Information Processing Societies Conference Proceedings* (Association for Computing Machinery, 1968), p. 1285.
Schwarz, C. V. and White, B. Y., Cogn. Instr. **23**(2), 165–205 (2005).
Selen, M., *Paper Presented at the APS April Meeting 2013* (APS, 2013).
Sherin, B. L, (University of California, Berkeley, 1996), see http://www.sesp.northwestern.edu/docs/dissertation/386770764374ec34c2565.pdf
Sherin, B. L., Cogn. Instr. **19**(4), 479–541 (2001).
Sherin, B. *et al.*, Interact. Learn. Environ. **3**(2), 91–118 (1993).
Sherwood, B. A. (2017). "A time line for VPython development," see https://brucesherwood.net/?p=136
Sherwood, B. A., Am. J. Phys. **39**(10), 1199–1202 (1971).
Smith, S. G. and Sherwood, B. A., Science **192**, 344–352 (1976).
Sokoloff, D. R. and Thornton, R. K., *Proceedings of ICUPE* edited by E. F. Redish and J. S. Rigden (AIP, 1997), Vol. 399, Issue 1061, pp. 1061–1074.
Sokoloff, D. R. *et al.*, Eur. J. Phys. **28**, S83–S94 (2007).
Solomon, C., *Educational Computer Environments for Children: A Reflection on Theories of Learning and Education* (MIT Press, Cambridge, Massachusetts, 1986).
Solvang, L. and Haglund, J., *Proceedings of EDULEARN18 Conference*, *July* (IATED, 2018), pp. 9667–9674, see https://www.diva-portal.org/s2013;mash/get/diva2:1249564/FULLTEXT01.pdf
Solvang, L. and Haglund, J., Phys. Educ. **56**(5), 055011 (2021).
Steinberg, R. N. *et al.*, AIP Conf. Proc. **399**(March 2013), 1075–1092 (1997).
Stephens, A. L. and Clement, J. J., Phys. Rev. Spec. Top. Phys. Educ. Res. **6**(2), 020122 (2010).
Strzys, M. P *et al.*, *Physics Education Research Conference Proceedings, 2018* (American Association of Physics Teachers, 2019).
Svensson, K. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010127 (2020).
Sweller, J., Cogn. Sci. **12**(2), 257–285 (1988).
Sweller, J., J. Educ. Psychol. **81**(4), 457–466 (1989).
Wolfe, A., Comput. Phys. **2**(3), 16 (1988).
Thees, M. *et al.*, Comput. Hum. Behav. **108**, 106316 (2020).
Thornton, R. K. and Sokoloff, D. R., Am. J. Phys. **58**(9), 858–867 (1990).
Thornton, R. K. and Sokoloff, D. R., AIP Conf. Proc. **399**, 1101–1118 (1997).
Thornton, R. K., Phys. Educ. **22**(4), 230 (1987).
Tinker, R. F., Phys. Teach. **19**(2), 94–105 (1981).
Tinker, R. F., *Advanced Educational Technologies for Mathematics and Science* (Springer, Berlin, 1993), pp. 91–113.
Tinker, R. F., *A History of Probeware* (2000). Accessed online http://makingsens.stanford.edu/resources.html
Treagust, D. F. *et al.*, *Multiple Representations in Physics Education*, edited by D. F. Treagust *et al.*, (Springer International Publishing, 2017), Vol. 10.

Triona, L. M. and Klahr, D., Cogn. Instr. **21**(2), 149–173 (2003).
Trowbridge, D. E. and McDermott, L. C., Am. J. Phys. **48**(12), 1020–1028 (1980).
Trowbridge, D. E. and McDermott, L. C., Am. J. Phys. **49**(3), 242–253 (1981).
Tufte, E. R., *The Visual Display of Quantitative Information* (Graphics Press, Cambridge, MA, 1983).
Tuminaro, J. and Redish, E. F., Phys. Rev. Spec. Top. Phys. Educ. Res. **3**(2), 020101 (2007).
Tversky, B., *Visualization in Science Education*, edited by J. K. Gilbert (Springer, 2005), pp. 29–42.
Uhden, O. *et al.*, Sci. Educ. **21**(4), 485–506 (2012).
van der Meij, J. and de Jong, T., Learn. Instr. **16**(3), 199–212 (2006).
Van Heuvelen, A. and Zou, X., Am. J. Phys. **69**(2), 184–194 (2001).
Van Heuvelen, A., Am. J. Phys. **59**(10), 891–897 (1991).
Van Heuvelen, A., Am. J. Phys. **69**(11), 1139–1146 (2001).
Van Heuvelen, A., AIP Conf. Proc. **399**(1997), 1119–1136 (1997).
Vliora, E. *et al.*, J. Open Distance Educ. Educ. Technol. **14**(2), 76–94 (2018).
Volkwyn, T. S. *et al.*, Des. Learn. **11**(1), 16–29 (2019).
Volkwyn, T. S. *et al.*, Learn. Res. Pract. **6**(1), 88–107 (2020a).
Volkwyn, T. S. *et al.*, *2017 Physics Education Research Conference Proceedings* (American Association of Physics Teachers, 2018), pp. 408–411.
Volkwyn, T. S. *et al.*, Eur. J. Phys. **41**(4), 045701 (2020b).
Walsh, T., Phys. Teach. **55**(5), 316–317 (2017).
Wee, L. K. and Ning, H. T., Phys. Educ. **49**(5), 493–499 (2014).
Wee, L. K., Phys. Educ. **47**(3), 301–308 (2012).
Wee, L. K. *et al.*, Phys. Educ. **47**(4), 448–455 (2012).
Wee, L. K. *et al.*, Phys. Educ. **50**(2), 189–196 (2015a).
Wee, L. K. *et al.*, Phys. Educ. **50**(4), 436–442 (2015b).
Wenning, C. J., J. Phys. Teach. Educ. Online **3**(1), 3–10 (2005).
White, B. Y. and Frederiksen, J. R., *Innovations in Science and Mathematics Education*, edited by M. J. Jacobson and R. B. Kozma (Routledge, 2000), p. 39.
White, B. Y., *New Directions in Educational Technology*, edited by E. Scanlon and T. O'Shea (Springer-Verlag, 1992), pp. 227–242.
White, B. Y., Cogn. Instr. **1**(1), 69–108 (1984).
White, B. Y., Cogn. Instr. **10**(1), 1–100 (1993).
White, B. Y. and Horwitz, P., *ThinkerTools: Enabling Children to Understand Physical Laws* (1987).
Whitelock, D. *et al.*, *Proceedings of EuroAIED* (Edicoes Colibri, 1996), see http://opensigle.inist.fr/handle/10068/410795
Wieman, C. E. *et al.*, Science **322**(5902), 682–683 (2008).
Wieman, C. E. *et al.*, Phys. Teach. **48**, 225–227 (2010).
Wilson, J. M. and Redish, E. F., Phys. Today **42**(1), 34–41 (1989).
Wilson, J. M., Phys. Teach. **32**(9), 518–523 (1994).
Wu, X. *et al.*, *2015 Physics Education Research Conference Proceedings* (American Association of Physics Teachers, 2015), pp. 383–386.
Yoon, S. A. *et al.*, Res. Sci. Technol. Educ. **36**(3), 261–281 (2018).
Zacharia, Z. C. and de Jong, T., Cogn. Instr. **32**(2), 101–158 (2014).
Zacharia, Z. C., J. Comput. Assist. Learn. **23**(2), 120–132 (2007).
Zietsman, A. I. and Hewson, P. W., J. Res. Sci. Teach. **23**(1), 27–39 (1986).
Zollman, D. A. and Fuller, R. G., Phys. Today **47**(4), 41–47 (1994).
Zollman, D. A., Phys. Teach. **22**(8), 514 (1984).
Zollman, D. A., Am. J. Phys. **64**(2), 114–119 (1996).
Zollman, D. *et al.*, J. Educ. Technol. Syst. **15**(3), 249–258 (1987).
Zu, T. *et al.*, *2017 Physics Education Research Conference Proceedings* (American Association of Physics Teachers, 2018), pp. 472–475.

SECTION V

PHYSICS EDUCATION RESEARCH: HISTORY, METHODOLOGIES, THEMES

Section Editor
David E. Meltzer

One might well ask, in a handbook devoted entirely to physics education research, what specific role would be played by a section whose title is "Physics Education Research"? How, after all, could its contents be distinguished from the contents of the rest of the handbook? The answer that has guided us in selecting the chapters for the section is that there are certain general methods and general perspectives that underlie, arguably, all work done in PER. That is, these general methods and perspectives inform, in some manner, nearly all investigations into the teaching and learning of physics. The first chapter in this section, by David Hammer, offers a personal view of the evolution of research methodologies in PER and the principles that guide that evolution. The second chapter by Joseph Taylor and Larry Hedges explores general methods that have been developed during recent decades for strengthening the process of conducting education research and expediting the accumulation of actionable knowledge on learning in general and physics learning in particular. The chapter by John Stewart, John Hansen, and Lin Ding provides a thorough overview of quantitative research methods that have been used in PER, including some that have only recently gained prominence in the field. Similarly, Valerie Otero, Danielle Harlow, and David Meltzer outline qualitative research methods, some of which have been used in the field since its earliest days, and others that have continued to develop up to the present. Finally, Jenaro Guisasola, Kristina Zuza, Paolo Sarriugarte, and Jaume Ametller describe general principles of applying physics education research to the design and assessment of classroom learning materials.

CHAPTER
22 THE NECESSARILY, WONDERFULLY UNSETTLED STATE OF METHODOLOGY IN PER: A REFLECTION

David Hammer

Hammer, D., "The necessarily, wonderfully unsettled state of methodology in PER: A reflection," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 22-1–22-12.

This chapter is more a reflection than a review. Other chapters in this volume present more systematic coverage of methodologies. My purpose is to consider the state of methodology in PER in the broadest sense to suggest that it remains unsettled, in contrast to what many of us might expect from our backgrounds in physics.

In keeping with this volume, I will focus on PER. It is important to recognize, however, that research on learning in physics arose from and lives within research on learning and instruction much more broadly (National Research Council, 2000), which itself sits within still larger pursuits to understand the mind and minds. What parts of that larger work are relevant depends significantly on what researchers see as "physics": a body of information, a set of reasoning skills, an evolving set of practices, and a pursuit to understand. Each raises connections to other work, from research on learning in mathematics, history, and other disciplines (National Research Council, 2012) to research on how people develop motor skills or retain arbitrary information.

I say that as a preface to emphasize that PER is part of and connects deeply to other areas. Many in PER see and promote it as a subfield of physics, for both substantive and political reasons. With respect to empirical methodology, my focus here, that positioning may be misleading. PER originated from and remains in rich interaction with other efforts to study human thinking and experience.

The first and most basic contribution of PER has been to show that it is possible: Phenomena of learning and teaching can be studied through evidence and reasoning. That was not obvious, and it strains longstanding structures and expectations in higher education: Institutional policies and

https://doi.org/10.1063/9780735425514_022

practices treat teaching and research as separate categories.[1] Physics education is a recent entrant into the set of areas humanity has begun learning how to study through deliberate, disciplined investigation.

Physicists have some sense of how that entry can happen. The tradition in Western curricula has been that physics began as an empirical science around 1600 in Europe with Galileo, but that was a re-introduction. There is clear historical evidence that it started at least 600 years earlier in the Middle East, with al-Ḥasan ibn al-Haytham (known as "Alhazen"):

> In a critical treatise, *Aporias Against Ptolemy*, [Ibn al-Haytham] asserts that "Truth is sought for itself"—but "the truths," he warns, "are immersed in uncertainties" and the scientific authorities (such as Ptolemy, whom he greatly respected) are "not immune from error…." Nor, he said, is human nature itself: "Therefore, the seeker after the truth is not one who studies the writings of the ancients and, following his natural disposition, puts his trust in them, but rather the one who suspects his faith in them and questions what he gathers from them, the one who submits to argument and demonstration, and not to the sayings of a human being whose nature is fraught with all kinds of imperfection and deficiency. Thus the duty of the man who investigates the writings of scientists, if learning the truth is his goal, is to make himself an enemy of all that he reads, and, applying his mind to the core and margins of its content, attack it from every side. He should also suspect himself as he performs his critical examination of it, so that he may avoid falling into either prejudice or leniency." (Sabra, 2003).

This is the earliest expression I have seen of core values we aspire to uphold in science. I take it as a starting point here: Methodology is born out of the expectation that anyone can be wrong or ignorant, including ourselves. It takes discipline of mind to "submit to argument and demonstration," rather than to accept authority or what seems obvious. The emergence of PER, like physics earlier, is in part an emergence of disciplined humility.

That is, physics education is yet another area in which we cannot entirely trust in our traditions or ourselves. What, then, can we do to learn? How can we come to know? How can we check what we believe; how can we be sure we're not deceiving ourselves or missing something? We devise methods for forming, assessing and refining knowledge. The added layer to consider is that methodology is also knowledge that reflects our imperfections and deficiencies. If we are "true seekers," we cannot simply study "the sayings of a human being" and put our trust in them to know how to learn.

In this moment, of course, I am drawing on the writings of an ancient, ibn al-Haytham (by way of Abdelhamid Sabra), and I should pause to be "the enemy" of what I read. It is, for one, sexist—"the duty of the man" and all—something I do not want to perpetuate. Last semester (Spring 2022) I worked out

---

[1] Henderson and Dancy (2007, 2011) found that "[m]ost faculty work in institutions where structures have been set up to work well with traditional instruction." (2011, p. 7); Corbo *et al.* (2016) argued that "the changes required for the systemic use of research-based teaching practices […] challenge existing norms and structures." My contention is similar, but with respect to the conduct of research on learning and instruction: It also challenges institutional norms and structures.

a gender-neutral edit of the quote to include in an assignment for my course (General Physics I). But my teaching assistants convinced me that editing would do different harm, of disguising a history that remains with us, and the quote went in as above.

That history, and helpful feedback from a colleague on my first draft, has me wonder also about the emphasis on "imperfection and deficiency." The message of human fallibility is important for *me* and others like me who are so often at risk of arrogance. There are more others who are often at risk of the opposite, living in a world that has not expected they have something to contribute (Prescod-Weinstein, 2020; and Barthelemy *et al.*, 2022). Re-reading the text, I can see it as empowering: True seekers question "the writings of the ancients." Or: "In questions of science, the authority of a thousand is not worth the humble reasoning of a single individual" (Galileo, 1632). Argument and demonstration can support ideas, including those that disrupt currently dominant views.

22.1 A REMINDER OF METHODOLOGY IN PHYSICS

Part of the challenge for many of us in PER is that we are involved in it during its formation, typically after being trained and enculturated in the much older discipline of physics. There we have had the privilege of centuries of progress, which like other privileges can be transparent. It will help to review, even briefly, some history of methodology in physics.

Time and again, there have been "revolutions" of mind (Kuhn, 1970) away from "truths" so obvious, or settled by authority, that no one was thinking to question them. Ibn al-Haytham led one, overturning settled ideas about light; Galileo another, about absolute rest and motion; Meitner another, about immutable elements. The revolutions often entailed showing some foundational aspect of thought to be an assumption—that objects have to be in contact to interact, that space and time are independent, that "elements of reality" (Einstein *et al.*, 1935) exist without observation, and so on and on (Holton and Brush, 2001).

Changes in how physicists conceptualize physical phenomena have come with changes in methodology. The shifts from thinking of light as a particle to light as a wave in a medium, then dropping the medium, then to its quantization as photons, all drew from and affected the experimental methodology. Those are examples of changes in *ontology*, that is, in the kinds of entities a model takes to exist.

On a simple, classical view of phenomena, throughout the 19th century, methods of empirical research depended on reproducibility: The same initial conditions lead to the same outcomes, within measurement errors. Quantum mechanics introduced an ontological randomness, inherent and irreducible, which required a change in expectations of reproducibility, to *distributions* of outcomes. There were changes in research methods that many physicists, including Einstein, thought meant the end of physics as a science. Chaos theory introduced an epistemological "randomness," as another challenge to expectations of reproducibility: Even assuming classical determinism, tiny differences in initial conditions can lead to radically different outcomes, in systems as simple as a dripping water faucet or a double-pendulum.

The main points here are that (1) the empirical methodologies of research are deeply entangled with the tacit or explicit theoretical models, ontologies and epistemologies of a discipline, and (2) the methodologies develop over time, sometimes in ways that earlier scholars could not have imagined. Over and over, the history of physics supports Ibn al-Haytham's insistence on human fallibility as well as the value of radically new thinking: Ideas that initially seem outlandish can take over the discipline.

In most areas of physics today, there are well-established theoretical foundations coupled with methods of research that have extensive track records of productivity. Those of us raised in those foundations and methods might forget that they were hundreds of years in formation, and that can make us impatient with what can seem like sloppy, undisciplined approaches to studies of learning and instruction. But disciplined study does not mean putting our trust in tradition. Methods we know from prior studies should inform how we approach research on learning, but we should not confuse adherence to authority with "rigor." Rigor should mean the ongoing pursuit of knowledge that is defensible upon close examination and argumentation.

Physics Education Research is new as a recognized pursuit, and it will remain new for our lifetimes. That gives us a different privilege, that of experiencing and contributing to its creation, which means—whether we enjoy it or not—our grappling with more uncertainties and ambiguities and possibilities than if our research were in (most areas of) physics. In the next section, I turn to an overview of that grappling.

22.2 METHODOLOGY IN PER

PER has involved an eclectic, disparate mix of methodologies, reflecting and entangled with an eclectic, disparate mix of aims, theoretical frameworks, as well as proto-theoretical assumptions that we might not notice. How we conduct our studies depends on and affects what we're after, functionally, and how we conceptualize its form. In what follows, I will sample from studies in PER, from work in the early years and from work in recent years.

22.2.1 The beginnings of PER

The first empirical paper of modern PER,[2] to my knowledge, was Reif *et al.* (1976), based on Larkin's dissertation and titled "Teaching general learning and problem-solving skills." It described two studies, one of the effects of teaching a "general learning skill" for learning "any new relation" (p. 212) and

---

[2] David Meltzer showed me research from earlier, "a few dozen research studies were published from 1910–1945" (Meltzer, 2015; see also Meltzer and Otero, 2015). The work I cite here, as the beginnings of modern PER, was evidently an independent re-emergence.

the other of the effects of teaching a "simple problem-solving strategy." The methods were a mix of qualitative and quantitative. Describing the study of general learning skills, the authors wrote,

> Our most important assessment method consisted of detailed observations of individual students. Such observations are essential to elucidate how students learn and to obtain the information needed for improving the teaching materials and the underlying models upon which they are based. (p. 214)

Describing their study of problem-solving strategies, the authors specified how they posed students problems, citing the methodology of protocol analysis from research on problem solving in cognitive science (Newell and Simon, 1972). Their protocols showed

> that many students in an introductory physics course approach problems in very haphazard and ineffective ways.... Thus, even when students know all the relevant facts and principles necessary for the solution of a problem, they may be unable to solve it because they lack any systematic strategy for guiding them to apply such facts and principles. (p. 216)

The quantitative parts of their study consisted of randomly dividing students to receive "special" or "ordinary" instruction and then testing the outcomes.

At roughly the same time, Laurence Viennot was working on her dissertation (Viennot, 1977), studying "spontaneous reasoning in elementary dynamics." She developed a pencil-and-paper questionnaire and administered it to several hundred students, mostly from Belgian and British universities. Today we would recognize her questions as valid probes of conceptual understanding, but at the time Viennot heard objections that they were "traps" for students. She argued that it was essential to pose questions different from those students had become familiar with in conventional teaching. In this regard, she was in line with widely accepted expectations of research, that to make a phenomenon apparent it is not always sufficient simply to observe. Researchers construct investigations, and that should not be different for research on learning (Viennot, 1977, p. 5).

The quantitative results—tallies of correct and incorrect answers—was the first evidence presented within PER of students not learning concepts in ways that instructors assumed. Qualitative analyses of students' written explanations showed several patterns of reasoning and multiple "notions" of force that students use, "depending on the question asked" (Viennot, 1979). Viennot drew both on cognitive psychology, mainly Piaget (1973), as well as on the history and philosophy of science, mainly Koyré (1966).

Trowbridge and McDermott (1980) followed shortly after with a study of students' reasoning in kinematics. They began with an "individual demonstration interview," which they described as "like the 'clinical interview' pioneered by Jean Piaget": "While the questioning follows a regular format, it allows for exploring any particular aspect of the student's thinking that may be of interest. Each interview lasts from 20 to 30 min and is audiotaped or occasionally videotaped. The dialog is transcribed and analyzed in detail."

They conducted over 300 of these interviews, initially drawing tasks from Piaget's research, and eventually tailoring their own speed comparison tasks that they used in pre- and post-course interviews. They identified a particular difficulty in students' confusion of speed with position, which enabled them to craft written questions they could include for students in course examinations.

Their analyses led them to conclude that "prior to instruction the student typically has a repertoire of procedures, vocabulary, associations, and analogies for interpreting motion in the real world," which they described as "a set of protoconcepts which antedate understandings of the concepts of kinematics" (p. 1027). They also found, just as Viennot argued in defense of her methods, that

> The ability to solve conventional problems on examinations or to pass the usual types of "mastery" tests does not always indicate conceptual understanding. Only certain types of questions can probe for the ability to resolve concepts from one another and to apply them to real situations. (p. 1028)

Right away, in the first few years of PER, there were multiple aims for research and methodologies. Reif, Larkin, and Brackett (1976) wanted to understand how to help students learn effective skills, with an expectation that these skills are independent of domain. They saw the target phenomena of reasoning as complex enough to require detailed observations for the study. They started from observations of where students began, and from those observations saw students as "haphazard," *lacking* in skills and strategies. Their experiment was to impart skills and strategies in the form of explicit, step-by-step instructions. Seeing these skills and strategies as domain general, their methods involved testing for outcomes in both physics and accounting.

Viennot's (1977, 1979) aim was to reveal and study phenomena of student reasoning that she also saw as complex, which also motivated methods of close, qualitative analyses of students' explanations for their answers. In other respects, her research was quite different: The phenomena she aimed to study concerned student conceptual reasoning about force; Reif *et al.* (1976) focused on how students learn. Viennot's analyses led her to see students as *having* multiple, mutually inconsistent notions of force they would apply depending on the question. They were not reasoning in ways physics instructors would hope, but Viennot saw substance and structure relevant for physics education, reflecting a constructivist epistemology.

Trowbridge and McDermott's (1980) methods started like Reif *et al.*'s (1976). They, too, saw the phenomena as needing a qualitative study, and they drew methodology from cognitive psychology. Like Viennot, they focused on conceptual understanding, and like her saw students' extant understanding as important to understand. While they emphasized "student difficulties," and designated them as the focus of later instructional intervention, they attributed to students protoconcepts important for their learning. Insight into those protoconcepts supported the researchers' designing specific, written probes.

It is striking in retrospect how these early studies were a mix of qualitative and quantitative methods, with the former more about initial exploration of the phenomena, and the latter to quantify some particular aspect. Methodology was significantly informed by much more general research on learning, from the cognitive sciences in the case of Reif *et al.* (1976), and it was significantly driven by computational models of mind. In line with emphases in that body of work, Reif *et al.* (1976) focused their investigations on learning and problem-solving skills. Viennot (1977) and Trowbridge and McDermott (1980) drew on Piagetian constructivism, which was not at the time influenced by computational modeling, and they focused on conceptual knowledge.

PER would go on to focus much more attention on conceptual understanding as the modal target of research, beginning mainly around phenomena of motion and force in the domain of Newtonian mechanics. These two early studies presented similar ontologies of conceptual knowledge, attributing to students multiple "notions" (Viennot, 1977) or "protoconcepts" (Trowbridge and McDermott, 1980). These ideas were similar to those diSessa (1979, 1982) was formulating, which he would later present as "knowledge in pieces" (diSessa, 1988).

Other studies presented a unitary ontology, attributing to students a "stable, alternative view" (Clement, 1982) or "a rich accumulation of interrelated ideas that constitute a personal system of common-sense beliefs" (Champagne *et al.*, 1980) that operates as something like a "paradigm" in competition with Newtonian mechanics. This ontology, or forms of it, became more common in PER for quite some time and was reflected in most accounts of misconceptions.

One reason for the prevalence of unitary ontologies may have been—may be—their methodological tractability: Stable properties are much easier to investigate. Unlike the other early studies I have described, Champagne *et al.* (1980) conducted their research by developing instruments to measure students' pre- and post-instruction conceptual knowledge, abilities for logical reasoning, and mathematical skills, which allowed statistical analyses of relationships.

They were drawing on disciplinary practices in psychology, not Piagetian clinical interviews but quantitative methods that resonate well with, and may have been influenced by, disciplinary practices in physics: "In physical science," Lord Kelvin wrote, "the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it." That stance has been widely adopted by psychologists, of quantifying qualities of mind and minds, including, significantly, through the development of instruments.

A great deal of PER has happened through the development of instruments (Ding, 2019), perhaps for this resonance between the disciplines of psychology and physics, a resonance of epistemology or esthetics. These methods support and are supported by conceptualizations of ontology and attributions of stable properties to individuals, such as conceptions, levels of cognitive development, attitudes, or beliefs. Qualitative methods, in contrast, more easily afford the conceptualization of situated dynamics, short-lived states of reasoning.

Along the way, the community has come to see a divide between qualitative and quantitative (Otero and Harlow, 2009; Ding and Liu, 2012; and Robertson *et al.*, 2018). It has also mattered that PER as a community, and as an enterprise, sits within and in close contact with the educational systems of K-12 and higher education, with its conceptualizations, values, and practices. Notions of efficient, objective assessment—such as in standardized tests—are powerfully attractive within those systems.

Some researchers developing instruments have cautioned against their use for individual attributions—to "use considerable care in applying the results of a limited probe such as our survey to a single student" (Redish *et al.*, 1998), noting "survey results for an individual student may be misleading," while still having considerable value as measures of the statistical ensemble. Redish *et al.* (1998) delivered that caution for having recognized the need: When an instrument produces a number for a student, a researcher or educator may be inclined to attribute it as a property.

22.2.2 A sampling of recent work

Since those beginnings, the goals and methods of PER have only expanded in variety and complexity. The study targets have gone beyond conceptual knowledge to include disciplinary practices of learning, attitudes, expectations, epistemologies, affect and emotions, and the effects of contexts and community (Docktor and Mestre, 2014). Theoretical frameworks have expanded as well, and diversified, with the dynamics growing ever more complex (Brown, 2014), drawing on and leading to new developments in methodology.

I refer readers to other chapters in this volume for more systematic reviews of current work (Guisasola *et al.*, 2023; Otero *et al.*, 2023; Stewart *et al.*, 2023; and Taylor and Hedges, 2023). Here, I mean only to call attention to the rich and fertile diversity of research taking place, from my own readings in the literature as well as from suggestions by colleagues.[3]

There are studies of student inquiry in "naturalistic" settings (if a physics course is part of nature!). Euler *et al.* (2019) focused close attention on 2.5 min of physics reasoning by a pair of introductory students working to make sense of binary-star dynamics, showing the students' advancing by dancing and gesturing, finding a range of ways of thinking to support physics sense-making. What Euler *et al.* presented generalizes, not necessarily as true of all students at all times, but in identifying and characterizing possible dynamics of student inquiry.

Other studies that have focused on close examinations of particular instances include Suárez (2020) studying how a group of bilingual students drew on different language resources in thinking about electricity in circuits; Odden and Russ (2019) finding affective and linguistic markers of *vexation* as initiating and sustaining inquiry; Sayre and Irving (2015) identifying markers of metacognition in students' brief, spontaneous interjections; and Kapon *et al.* (2018) examining student inquiry in two

---

[3] I am grateful for suggestions by Amy Robertson and an anonymous reviewer.

settings to explore tensions and connections between goals of authenticity to physics and of personal relevance to students.

There are also studies based on interviews. One example is by Moshfeghyeganeh and Hazari (2021), who set out to understand a phenomenon evident in large-scale patterns: Women are much better represented as physics students in Iran than in the United States, and the pattern is also observed across other Muslim Majority (MM) countries. Why is that? The researchers interviewed seven faculty members in physics, all women who emigrated from MM countries, asking them to recount their experiences and to reflect on expectations in their communities. Their work led to several possible hypotheses, such as that "expressions of femininity emphasize modesty" in MM countries, "rather than physical attractiveness," are a closer match to values in the physics community.

Surveys of course remain a prominent and important approach to research at larger scales (Madsen *et al.*, 2017). For one recent example, Deslauriers *et al.* (2019) used a randomized experimental approach to compare "passive lectures with active learning." They used two instruments, one a test of conceptual understanding and the other a Likert scale survey of students' agreement or disagreement with the statement "I feel like I learned a great deal from this class." The findings showed that students' feelings of learning were anticorrelated with their scores on the conceptual test.

PER began with studies that worked across qualitative and quantitative studies, and it seems like a valuable heuristic for the field: Look for evidence in multiple forms, as a community if not as individual researchers. Hypotheses from Moshfeghyeganeh and Hazari (2021) could inform the development of larger-scale surveys; aggregate correlations in Deslauriers *et al.* (2019) could motivate focused small-N study of the phenomena. Little *et al.* (2019) took a novel approach to studying an established idea, Dweck's (2013) construct of *mindset*, by coding interview dialogue and finding nuance they would have missed using the usual Likert surveys. John & Allie designed a series of studies specifically to connect findings across methods, starting with a multiple-choice instrument (John and Allie, 2017a), then free writing responses (John and Allie, 2017b), and finally naturalistic "micro-episodes" (John and Allie, 2019), all to study the contextuality and complexity of student reasoning around DC circuits.

In PER as in other fields, new technologies afford new methodologies; the possibility of video recording has had a powerful effect on research in the learning sciences (Derry *et al.*, 2010). More recent developments have scholars using eye-tracking (Ibrahim and Ding, 2021; and Wu and Liu, 2021) and imaging of brain activity (Allaire-Duquette *et al.*, 2021), connecting the evidence from these sensors to findings from other modes of research. PER has begun to draw on methods of machine learning and data science (Yang *et al.*, 2020; and Aiken *et al.*, 2021), including to support qualitative data analysis (Sherin *et al.*, 2018; and Çınar *et al.*, 2020). These approaches may provide new ways to bridge research across the qualitative/particular and the quantitative/general. (On the last approach I am uneasy that the dynamics of our communities will rush to put such tools into practical implementation—automatic grading in courses. Here, I am expressing interest in their use for research.)

22.3 LOOKING FORWARD

PER remains a young area of study. We have made progress, no question, adapting methods of research including interviews, observations, and surveys that will continue to shape our work. But the precise forms of these methods, and our methodologies more generally remain significantly unsettled in tandem with our theoretical foundations. This unsettled state might be discomforting, especially for those of us who were first trained in physics, where we had the benefit of working within long established frameworks.

Of course, we will and should draw on ideas from other disciplines and areas of research. Physics and psychology have been our two leading source fields; in recent years, PER has looked to sociology (e.g., Goffman, 1974) and to critical race and gender studies (Traxler *et al.*, 2016; and Rodriguez *et al.*, 2022). My colleagues and I recently argued for drawing from research in ecology (Hammer *et al.*, 2018).[4] The subfield of community ecology in particular has also struggled with how to handle the difficulties of complexity and idiosyncrasy in the phenomena they study. But there may well be more—we are, after all, studying *humans*. (Perhaps we should see PER as a subfield, not of physics, but of biology?)

Wherever we get our ideas, we should take care not to treat them as dogma. It is easy and tempting to settle back on authority, on "what everyone knows," on what seems clear and obvious. In many areas of physics, there are genuinely well-established methodologies, but physics as an empirical discipline has been around for at least 1000 years. And the moral from studying physics is that even the most foundational conceptualizations can change. The empirical study of learning and teaching in physics has only been around for decades: There should be no illusion that there are permanent "gold standards" of empirical scholarship (Cartwright, 2007).

My main objective in this essay has been to argue that we work as "true seekers," remembering along the way to "suspect [ourselves] as imperfect and deficient" (Sabra, 2003, p. 54). That means both staying humble, in our own scholarship and in assessing others', and embracing the possibility of novelty. It seems to be an occupational hazard for scholars, investing deeply into a point of view, and committing to it firmly, that we (myself certainly included!) can find ourselves policing the community for adherence. If we're doing that in our scholarship, as reviewers or editors or grant panelists, or even simply in positions of influence in the community, we can end up preventing new ideas from getting explored, considered, published, or funded.

That is not to say that we should let anything go. To the contrary, we should continue to question each other and ourselves, ancient (1980s?) and recent. My colleague Leema Berland and I published a critique to challenge some common and widely accepted practices in qualitative research (Hammer and Berland, 2014). In recent years for research quite broadly, quantitative methods have been challenged for the problems of p-hacking. But we should take care to construct arguments and evidence, open to the possibility of something different. The unsettled state means we can and should expect, welcome, and

---

[4] I am grateful to Julia Svoboda Gouvea for seeing that connection.

engage with diverse ideas, including those that depart from "truths" that seem like obvious common sense. Novel methods will need novel consideration, and PER is full of novel methods.[5]

I have argued that PER needs to be in an unsettled state as a new field taking on very difficult, very complex problems. I am also suggesting we see it as a wonderful privilege to be working in PER when the field is so new and dynamic and evolving. That things are unsettled allows for scholarly invention and imagination; there remains lots of unexplored intellectual terrain.

Teaching introductory physics, I try to convince students to embrace confusion and uncertainty. I tell them that physicists are professional learners and that being confused for a physicist is like breathing hard for a runner: It's what you're supposed to feel. It is difficult: They have grown up and still live in an educational system filled with messages that confusion and uncertainty are bad things to avoid. I was motivated to write this essay for my sense that we too, the PER community, feel systemic pressures to present clear, simple findings, to be too sure too soon. I hope this essay might help in some small way.

ACKNOWLEDGMENTS

First, I thank Paula Heron and Fatih Taşar for their gracious invitation to suggest a topic and for accepting this one. I am most grateful to David Meltzer, Amy Robertson, Roger Tobin, and two anonymous reviewers for their helpful and critical feedback.

REFERENCES

Aiken, J. M. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(2), 020104 (2021).
Allaire-Duquette, G. *et al.*, NPJ Sci Learn **6**(1), 11 (2021).
Barthelemy, R. S. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 010124 (2022).
Brown, D. E., Sci. Educ. **23**(7), 1463–1483 (2014).
Cartwright, N., BioSocieties **2**(1), 11–20 (2007).
Champagne, A. B. *et al.*, Am. J. Phys. **48**, 1074–1079 (1980).
Çınar, A. *et al.*, Educ. Inf. Technol. **25**(5), 3821–3844 (2020).
Clement, J., Am. J. Phys. **50**, 66 (1982).
Corbo, J. C. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(1), 010113 (2016).
Derry, S. J. *et al.*, J. Learn. Sci. **19**(1), 3–53 (2010).
Deslauriers, L. *et al.*, Proc. Natl. Acad. Sci. U.S.A. **116**(39), 19251–19257 (2019).
Ding, L. and Liu, X., *Getting Started in PER, Reviews in PER Vol. 2*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2012), Vol. 2.
Ding, L., Phys. Rev. Phys. Educ. Res. **15**(2), 020101 (2019).
diSessa, A. A., *Constructivism in the Computer Age*, edited by G. Forman and P. Putall (Erlbaum, Hillsdale, NJ, 1988), pp. 49–70.
diSessa, A. A., *Cognitive Process Instruction* (Franklin Institute Press, Philadelphia, 1979).
diSessa, A. A., Cogn. Sci. **6**, 37–75 (1982).

---

[5] During my work on this chapter, Robertson and Hairston (2022) published a novel method for analyzing the "whiteness" inherent in an "ordinary" interaction among students in introductory physics. It drew unscholarly, intolerant reactions, dismissing the work without genuine consideration, a striking case in point. Henderson and Thoennessen (2022) responded with an editorial, arguing for the importance of "thoughtful presentation and debate of ideas in a welcoming environment" and "open discussion of sometimes controversial ideas."

Docktor, J. L. and Mestre, J. P., Phys. Rev. Spec. Top. Phys. Educ. Res. **10**(2), 020119 (2014).

Dweck, C. S., *Self-theories: Their Role in Motivation, Personality, and Development* (Psychology Press, 2013).

Einstein, A. *et al.*, Phys. Rev. **47**(10), 777 (1935).

Euler, E. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(1), 010134 (2019).

Goffman, E., *Frame Analysis: An Essay on the Organization of Experience* (Harvard University Press, Cambridge, Mass, 1974).

Guisasola, J. *et al.*, in *International Handbook of Physics Education Research: Special Topics*, Chap. 6 (AIP Publishing, 2023).

Hammer, D. and Berland, L. K., J. Learn. Sci. **23**, 37–46 (2014).

Hammer, D. *et al.*, Infanc. Aprendiz. **41**(4), 625–673 (2018).

Henderson, C. and Dancy, M. H., Phys. Rev. Spec. Top. **3**(2), 020102 (2007).

Henderson, C. and Thoennessen, M., Phys. Rev. Phys. Educ. Res. **18**(1), 010001 (2022).

Henderson, C. *et al.*, paper presented at the Invited paper for the National Academy of Engineering, Center for the Advancement of Engineering Education Forum, Impact and Diffusion of Transformative Engineering Education Innovations, 2011.

Holton, G. J. and Brush, S. G., *Physics, the Human Adventure: From Copernicus to Einstein and Beyond*, 3rd ed. (Rutgers University Press, New Brunswick, NJ, 2001).

Ibrahim, B. and Ding, L., Phys. Rev. Phys. Educ. Res. **17**(1), 010126 (2021).

John, I. and Allie, S., Eur. J. Phys. **38**(1), 015701 (2017a).

John, I. and Allie, S., Eur. J. Phys. **38**(1), 015702 (2017b).

John, I. and Allie, S., Eur. J. Phys. **40**(5), 055704 (2019).

Kapon, S. *et al.*, Sci. Educ. **102**(5), 1077–1106 (2018).

Koyré, A., *Études Galiléennes* (Hermann Paris, 1966), Vol. 15.

Kuhn, T. S., *The Structure of Scientific Revolutions* (University of Chicago Press, Chicago, 1970).

Little, A. J. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(1), 010127 (2019).

Madsen, A. *et al.*, Am. J. Phys. **85**(4), 245–264 (2017).

Meltzer, D. E., paper presented at the 2015 Frontiers and Foundations in Physics Education Research, Bar Harbor, ME, 2015. http://physicseducation.net/talks/FFPER_2015_final.pdf.

Meltzer, D. E. and Otero, V. K., Am. J. Phys. **83**(5), 447–458 (2015).

Moshfeghyeganeh, S. and Hazari, Z., Phys. Rev. Phys. Educ. Res. **17**(1), 010114 (2021).

National Research Council, *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* (The National Academies Press, Washington, DC, 2012).

National Research Council, *How People Learn: Brain, Mind, Experience, and School: Expanded Edition* (The National Academies Press, Washington, DC, 2000).

Newell, A. and Simon, H. A., *Human Problem Solving* (Prentice-hall Englewood Cliffs, NJ, 1972), Vol. 104.

Odden, T. O. B. and Russ, R. S., Int. J. Sci. Educ. **41**(8), 1052–1070 (2019).

Otero, V. K. and Harlow, D. B., *Getting Started in PER, Reviews in PER Vol. 2*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2009), Vol. 2.

Otero, V. K. *et al.*, in *International Handbook of Physics Education Research: Special Topics*, Chap. 25 (AIP Publishing, 2023).

Piaget, J., *La Formation de la Notion de Force* (Presses universitaires de France, 1973), Vol. 29.

Prescod-Weinstein, C., Signs: J. Women Cult. Soc. **45**(2), 421–447 (2020).

Redish, E. F. *et al.*, Am. J. Phys. **66**(3), 212–224 (1998).

Reif, F. *et al.*, Am. J. Phys. **44**(3), 212–217 (1976).

Robertson, A. D. and Hairston, W. T., Phys. Rev. Phys. Educ. Res. **18**(1), 010119 (2022).

Robertson, A. D. *et al.*, *Getting Started in PER, Reviews in PER Vol. 2*, edited by C. Henderson and K. A. Harper (American Association of Physics Teachers, College Park, MD, 2018), Vol. 2.

Rodriguez, M. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 013101 (2022).

Sabra, A. I. (2003). Ibn al-Haytham: Brief life of an Arab mathematician. Harvard Magazine, September–October, pp. 54–55.

Sayre, E. C. and Irving, P. W., Phys. Rev. Spec. Top. **11**(2), 020121 (2015).

Sherin, B. *et al.*, *Learning Analytics in Support of Qualitative Analysis* (International Society of the Learning Sciences, Inc., 2018).

Stewart, J. *et al.*, in *International Handbook of Physics Education Research: Special Topics*, Chap. 24 (AIP Publishing, 2023).

Suarez, E., Sci. Educ. **104**(5), 791–826 (2020).

Taylor, J. A. and Hedges, L., in *International Handbook of Physics Education Research: Special Topics*, Chap. 23 (AIP Publishing, 2023).

Traxler, A. L. *et al.*, Phys. Rev. Phys. Educ. Res. **12**(2), 020114 (2016).

Trowbridge, D. E. and McDermott, L. C., Am. J. Phys. **48**(3), 1020–1028 (1980).

Viennot, L., Eur. J. Sci. Educ. **1**(2), 205–221 (1979).

Viennot, L., *Le Raisonnement Spontane en Dynamique Elementaire* (Universite Paris VII, 1977).

Wu, C.-J. and Liu, C.-Y., Phys. Rev. Phys. Educ. Res. **17**(1), 011205 (2021).

Yang, J. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020130 (2020).

CHAPTER

23

TOWARD MORE RAPID ACCUMULATION OF KNOWLEDGE ABOUT WHAT WORKS IN PHYSICS EDUCATION: THE ROLE OF REPLICATION, REPORTING PRACTICES, AND META-ANALYSIS

Joseph A. Taylor and Larry V. Hedges

Taylor, J. A. and Hedges, L. V., "Toward more rapid accumulation of knowledge about what works in physics education: The role of replication, reporting practices, and meta-analysis," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 23-1–23-20.

23.1 CRITICISMS AROUND KNOWLEDGE ACCUMULATION FOR WHAT WORKS IN EDUCATION

In the late 1970s, there was considerable epistemological pessimism in education research. A great deal of educational research had been done in the 1960s and 1970s, but the failure of results to replicate led to a sense that a systematic base of knowledge for improving education was not accumulating. Some of the most distinguished scholars in the field questioned whether generalizable knowledge was even possible and advocated abandoning the search for such generalizations in education (e.g., Cronbach, 1975, 1982).

https://doi.org/10.1063/9780735425514_023

One response to this pessimism emerged from the systematic study of how research evidence accumulated (or did not) in education. Scholars in education and psychology began to examine the process by which reviews of research reached their conclusions and found them highly problematic. One strand of this work examined the methods used in published research reviews. These scholars drew an analogy between conceptual steps in primary (original) studies and the accepted methodological standards for conducting primary studies. They reasoned that problem formulation, data collection, data evaluation, and data analysis and interpretation are part of the methodology of both primary research and research reviews. They argued that research reviews should be just as transparent and rigorous as primary research and that similar methodological standards should apply to both (see, e.g., Cooper, 1982). Researchers analyzing actual reviews in education identified very serious shortcomings in virtually all reviews examined (see, e.g., Jackson, 1980).

Among the shortcomings identified was the use of qualitative methods for evaluating and synthesizing the results of quantitative studies. Some scholars showed how seemingly sensible qualitative synthesis methods were often highly misleading (see, e.g., Hedges and Olkin, 1985). Others focused on how to combine evidence across studies in more valid ways, creating the specialty in statistics called meta-analysis (see, e.g., Glass *et al.*, 1981; and Hedges and Olkin, 1985). Interestingly, the same progression of identifying problems in reviews of research and the emergence of interest in rigorous systematic reviews using meta-analysis also occurred in medicine a few years later than in education (see, e.g., Kass, 1981; and Mulrow, 1987).

During the 1980s and 1990s, rigorous research reviews using meta-analysis became more common in education. While clear evidence of progress in some areas of education research mounted, other areas of education research did not seem to advance. Many observers perceived the field of education research to be extremely weak. One sympathetic observer, the distinguished historian of education Carl Kaestle, wrote at the time that education research had an "awful reputation" both inside and outside the academy (Kaestle, 1993).

Toward the end of the 1990s, many observers noted that the U.S. economy had become a service- and knowledge-driven economy in which high levels of literacy and numeracy were necessary not just for the individual but also for national prosperity. A consensus emerged that rigorous scientific evidence was required to meet the challenge facing the nation in creating the kind of highly skilled workforce that was needed (see, e.g., Packer, 1997; Murnane and Levy, 1996; and Shavelson and Towne, 2002).

At this point there was not a complete consensus about how to think about the accumulation of knowledge in educational research, but a National Research Council study appeared in 2002 that offered important insights (Shavelson and Towne, 2002). Their analysis demonstrated that there were areas of strength in education research that *had* made important progress. They also argued that progress was not always continuous and unidirectional and that inevitable mistakes were sometimes the keys to later progress. They insisted that even though there was no simple formula for scientific progress, the essential elements of proposing theories, and evaluating those theories, both logically and empirically, are essential elements of how science makes progress and how it corrects mistakes.

Federal policy makers looked to medical research to find a better model for the institutional support of educational research. The National Institutes of Health (NIH) traces its origins to the founding of the Marine Hospital Service in 1887, but its present form and functions stem from the 1949 reorganization and expansion that gave it its present name. That was the beginning of a sea change in medical research. Coincidentally, the year before was the date of the first modern randomized trial, the streptomycin trial (Medical Research Council, 1948). Randomized clinical trials quickly became the gold standard for rigorous tests of causal effects in medicine at NIH and worldwide. Rigorous methods for systematic reviews and meta-analyses of trials in the 1990s soon led to the emergence of evidence-based medicine in which policies and practices are as informed as possible by actual research evidence.

This model migrated to education and was embodied by the Education Sciences Reform Act (U.S. Congress, 2002) which created the U.S. Institute of Education Sciences (IES, 2022a) as the research arm of the U.S. Department of Education. The IES included a federal statistical agency (the National Center for Education Statistics: NCES), a Center that largely funded contract research (the National Center for Educational Evaluation and Regional Assistance: NCEE), and two centers that funded field-initiated research: the National Center for Education Research (NCER), and the National Center for Special Education Research (NCSER). The latter two centers were modeled self-consciously upon the research funding programs at the National Institutes of Health, with high methodological expectations, standing review panels for evaluating research proposals, and high standards for rigorous peer review.

The IES research portfolio included funding for a phased set of research goals: exploration, development of interventions, efficacy testing, and effectiveness evaluation (as well as funding for replications, methodology and measurement research). It also included funding for systematic reviews of rigorous research and dissemination of those reviews through its What Works Clearinghouse (WWC: IES, 2022b). Perhaps the most dramatic innovations were the IES emphasis on randomized trials for efficacy and effectiveness studies and the WWC for dissemination of this work. For example, since 2002, IES' NCER and NCSER has funded approximately 800 grants containing over 900 planned randomized trials (with multiple group designs) in education (E. Albro, NCER Commissioner, personal communication, February 22nd, 2022). Further, in that same timeframe, projects funded by IES' NCEE produced another 80 randomized trials (E. Albro, NCER Commissioner, personal communication, 22 February, 2022). IES has adopted the medical model that randomized trials provide the most trustworthy evidence about cause and effect of educational interventions, products, or services—the gold standard.

These aspects of the IES program have been copied by other countries, most notably the United Kingdom, whose Educational Endowment Foundation (Educational Endowment Fund, 2022) has funded well over 150 randomized trials (Edovald and Nevill, 2021) and has a dissemination mechanism called the EEF Toolkit to provide evidence about the results of studies in education. In Denmark, Germany, Norway, and Sweden, there are evidence synthesis groups working on education and social welfare (e.g., Germany's Zentrum für internationale Bildungsvergleichsstudien; ZIB, 2022). These groups come together with others worldwide as part of the Campbell Collaboration (Campbell Collaboration, 2022), and ***i3e*** (International Initiative for Impact Evaluation, 2022), where the latter has funded 338 randomized trials of interventions (as of January 2022) as well as syntheses of work on development in middle and low-income countries. Other countries in Europe and elsewhere seem to be emulating these efforts to provide the basis for evidence-based policy and practice in both education and social services.

23.1.1 Parallel trends in physics education

Similarly, in physics education research, causal effect studies (e.g., experiments and quasi-experiments) are well-represented, being identified by Ding (2019) as one of the three primary genres of physics education research (PER). Further, there is some evidence that researchers studying interventions in the physical sciences have been more apt during this time period to use randomized designs than have researchers in other science education disciplines. The authors conducted a secondary analysis of publicly-available data (Kowalski and Taylor, 2019) from the Taylor *et al.* (2018) meta-analysis of effects from science education interventions, an internationally focused synthesis of 292 intervention effects for school-aged students (ages 5–18). In this analysis, it was observed that 99% (135/137) of studies of physical science interventions (physical science, physics, or chemistry) used randomized designs, compared to 69% (107/155) for interventions in other science education disciplines (biology, earth science, multidisciplinary science). However, smaller, more narrowly focused reviews of PER note a more modest proportion of randomized trials (Taasoobshirazi and Carr, 2008; and Uzunboylu and Aşıksoy, 2014).

In summarizing the characteristics of recent PER, it is helpful to discuss study design characteristics in terms of their implications for the validity of the respective studies. For the purpose of this chapter, the authors draw upon the validity typology of Shadish *et al.* (2002). In their typology, there are four types of validity, defined as follows: (a) *statistical conclusion validity* is the validity of inferences about the covariation between treatment and outcome, (b) *internal validity* is the validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship from A to B as those variables were manipulated or measured, (c) *construct validity* is the validity of inferences about the higher order constructs that represent sampling particulars, and (d) *external validity* is the validity of inferences about whether the cause–effect relationship holds over variation in persons, settings, treatment variables, and measurement variables (p. 38).

With regard to internal validity, a key consideration is mitigating *selection* bias where systematic differences in study participant characteristics exist across treatment conditions and these differences could produce the observed effect. In expectation, the random assignment used in RCTs balances participant characteristics across treatment conditions.

While a shift toward more rigorous impact designs (e.g., RCTs) has likely increased the internal validity of intervention studies in PER, this is just one step toward more rapid accumulation of knowledge about the effectiveness of physics education interventions. In the following sections, the authors describe important further measures that would move the field toward this goal.

23.1.2 The state of intervention research in physics education

In 2014, Fraser *et al.* reflected on the state of PER as a scientific field. From their assessment, they charged physics education researchers to either continue or begin to view PER as a science whose inherent studies should be subject to rigorous evaluation. Further, Fraser and colleagues strongly stated the need for PER to include more replications of studies that examine the effects of physics education interventions and subsequent meta-analyses of those effects.

While only a few modern meta-analyses exist in PER, meta-analyses in education now have a history of common use dating back to the 1980s (Hedges and Cooper, 2009). This movement toward modern meta-analysis was initiated in education by the landmark paper of Gene Glass (Glass, 1976), supported by advances in meta-analytic methods (Hedges and Olkin, 1985), and was preceded by a similar movement in medicine beginning in the mid-1970s (O'rourke, 2007).

In the language of Shadish *et al.* (2002), this call for replication and meta-analyses, where effect size heterogeneity can be quantified, corresponds to a call for better assessments of external validity. Specifically, the availability of multiple impact study replications and modest estimates of effect size heterogeneity would jointly address external validity concerns that treatment effects differ significantly across study samples, intervention variations, outcome measures, or study settings. Similarly, in the spirit of more generalizable meta-analytic results, Fraser *et al.* also charged primary study authors to disseminate all effects: negative, null, and positive. The detrimental effects of publication bias on negative or null effects are discussed later in this chapter.

The PER literature is not entirely void of syntheses (e.g., Taasoobshirazi and Carr, 2008; Ruiz-Primo *et al.*, 2011; and Docktor and Mestre, 2014). For example, Docktor and Mestre (2014) conducted a non-statistical synthesis of physics education research in the areas of conceptual understanding, problem solving, curriculum and instruction, assessment, cognitive psychology, and attitudes and beliefs about teaching and learning. Taasoobshirazi and Carr (2008) contributed a similar study on the topic of context-based physics. This kind of work is useful to the field but has some limitations. The nemesis of non-statistical syntheses is that it can be difficult to draw conclusions from them as they tend to weight all studies identically and the intervention effects are often characterized in a categorical way (e.g., positive or negative) without consideration for the numeric magnitude of those effects. Statistical approaches to synthesizing effects, such as fixed or random effects meta-analyses (Hedges and Olkin, 1985), address both of these limitations and the authors advocate this statistical approach in future PER studies. Fixed effect meta-analyses assume that there is one *true* effect size and that the only source of variation in effect size estimates is due to a sampling error (i.e., from different participants across trials), whereas random effects meta-analyses assume two sources of error: sampling error and differences in the *true* effects across interventions. Because random effects meta-analyses quantify the differences in true effects across studies, they provide information about whether the findings of studies are largely consistent or wildly disparate.

With regard to statistical syntheses, a few large-scale meta-analyses of science education interventions have been conducted in the last decade (e.g., Furtak *et al.*, 2012; Slavin *et al.*, 2014; Cheung *et al.*, 2017; and Taylor *et al.*, 2018). However, few have been specific to physics education and those that exist tend to be relatively small in scale (e.g., Madsen *et al.*, 2015). Physics education research would benefit from an increase in future syntheses, such as those that examine the effects of peer tutoring or group problem solving. These syntheses might expand their examination of treatment effects to contemporary outcomes of interest, such as those in the social/emotional domain (e.g., sense of community and belonging).

The authors acknowledge that perhaps the relatively small number of physics education meta-analyses could be tied to a paucity of studies available to be synthesized. It is likely that this deficiency in usable studies is tied to the replication crisis in education broadly (Makel and Plucker, 2014; and Hedges, 2018), in science education specifically (Taylor *et al.*, 2016), and to a version of outcome reporting bias where there is a measurable suppression of small or non-significant effects from being disseminated in the literature (Pigott *et al.*, 2013). The remainder of the chapter will describe considerations for how to facilitate more and better populated meta-analyses by promoting key study reporting practices that will facilitate an increased number of planned replications in physics education.

23.2 REPLICATION AND ITS CONTRIBUTION TO KNOWLEDGE ACCUMULATION

The concept of replication is central to the logic and rhetoric of science. The principle that scientific studies can be replicated by other scientists is part of the logic that science is self-correcting because attempted replications will identify findings that cannot be replicated and are possibly incorrect (see, e.g., McNutt, 2014). It is therefore not surprising that failures to replicate in the biomedical sciences got considerable attention (Ioannidis, 2005; and Perrin, 2014) and led to consternation in those fields (see, e.g., Collins and Tabek, 2014), as did similar findings in psychology (e.g., Open Science Collaborative, 2016), economics (e.g., Camerer *et al.*, 2016), and other areas of science (see, e.g., Bollen *et al.*, 2015; and Baker, 2016). At the same time, the authors acknowledge that some education researchers consider education too complex for studies to establish confident causal inferences and/or to expect effects to replicate across studies. The authors agree that education is very complex but suggest that strong causal inference is possible with rigorous designs, statistical controls, and consistent, high-fidelity implementation across trials. Whether studies providing strong causal inference can be replicated remains an empirical question.

23.2.1 Reproducibility and replication

An important distinction is that between reproducibility and replicability. Reproducibility concerns whether another investigator can obtain the same results (even literally identical numerical answers) when given the first investigator's research report and their data (and possibly the computer code they used to analyze the data). Replicability concerns whether another investigator can obtain the same results when they obtain their own (new) data by attempting to repeat the study that was carried out by the first investigator. A key difference between reproducibility and replicability is that the former involves whether two investigators can obtain the same answers when given the same data, but replicability involves whether two investigators can obtain the same answers from two different datasets. Replicability is more demanding than reproducibility and has been called "the ultimate standard by which scientific claims are judged" (Peng, 2011, p. 1226). Because "getting the same answer" is to some degree a matter of interpretation, and differences between two datasets may be induced by sampling variability, replicability is generally a more ambiguous concept than reproducibility.

23.2.2 Programs of replication in science

Systematic replication and synthesis of replicated studies have long been part of research in the physical sciences. One tradition of such work has focused on the so-called fundamental constants of mathematical physics. A seminal methodological paper on the joint estimation of these constants was published by Birge (1932) providing the basic statistical methods that are still used in combining evidence across studies to obtain the reference values in wide scientific use by the Committee on Data for Science and Technology (CODATA) in their periodic redetermination of the fundamental physical constants. Figure 23.1 (from the 2014 CODATA compilation, see Mohr *et al.*, 2016) illustrates the data from the individual studies (and their uncertainties) along with the synthesized values from 2010 and 2014. The process they use is essentially a form of meta-analysis. There are also many more such examples in physics and chemistry. The Particle Data Group (see Rosenfeld, 1975 or Olive *et al.*, 2014), housed jointly at UC Berkeley and CERN, has produced compendia of systematic reviews of research in particle physics every two years for over half a century. Other empirical series of measurements of physical quantities, including experimental measurements of the speed of light since 1870, are given in Hedges (2019).

A broader effort to review, synthesize, and disseminate standard reference data on physical, chemical, or biological properties of substances was created by the 1968 Standard Reference Data Act. It is notable that, to implement this act, critical evaluation standards were developed that required not only that "fundamental science and widely accepted standard operating procedures" be used for data collection, but also that "calculated and experimental data have been *quantitatively* compared" (the emphasis on quantitative comparison is directly from the Reference Data Act, not from the chapter authors). This standard effectively states that to be approved as refence values, there must be replication and the results of those replications must be compared statistically.

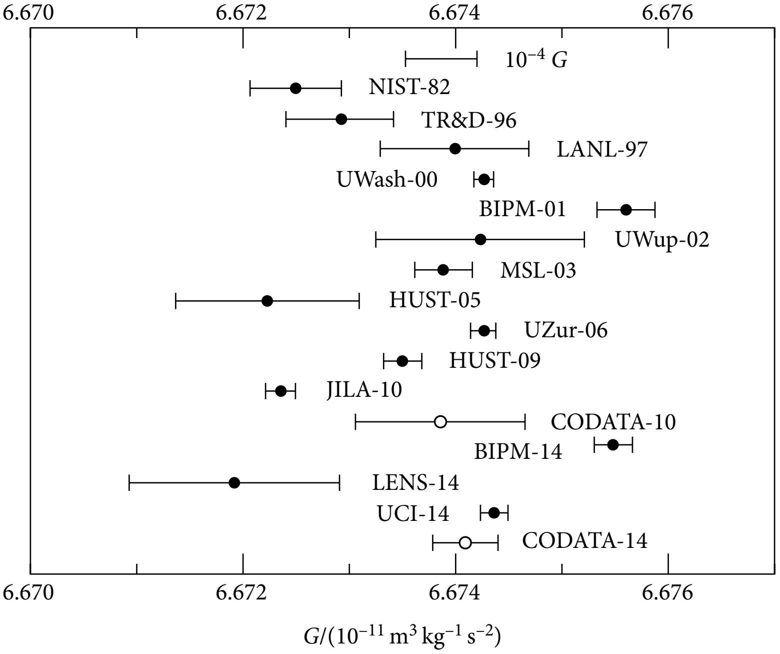


FIG. 23.1

Estimates of the universal gravitation constant from replicated experiments used in the 2014 determination of reference values (Mohr *et al.*, 2016).

23.2.3 Direct and conceptual replications

Within the concept of replicability, another important distinction is often made: between direct and conceptual replication (see Schmidt, 2009). *Direct replication* occurs when a replication study attempts to to keep research methods and procedures as similar as possible to those in the study to be replicated. That is, they seek to hold constant all the conditions and variables that are believed to possibly influence the outcome of the study. The joint NSF-IES Standards on Reproducibility and Replication (2018) somewhat overstates the case by arguing that "the goal of direct replication studies is to test whether the results of the previous study were due to error or chance" (p. 2). There is evidence that problems in replicating previous studies are often due to incomplete understanding of which aspects of methods or procedures are important in determining outcomes and which are not (see, Collins, 1992).

In contrast to direct replication, *conceptual replication* attempts to determine whether results are similar if the previous study's methods are systematically varied. One might argue that conceptual replication is a kind of generalization study, or in the words of Shadish *et al.*, 2002, a study of external validity. The distinction between direct and conceptual replication is not always completely clear because it is not always clear what methods or procedure matter in determining the results of a study. Experiments providing values of physical constants are often a combination of direct and conceptual replications as new experimental approaches provide different and typically more accurate estimates. However, not all new methods turn out to be improvements, but scientific progress sometimes results when a method or procedure is varied, which is not expected to change results but does, producing an anomaly that becomes the focus of new research.

23.2.4 What does it mean for studies to replicate?

There is no reason to think that any empirical study would yield *exactly* the same results if it could be repeated in exactly the same way. Statistical methods (e.g., the standard error of estimate) attempt to quantify the magnitude of the differences that might be expected among studies that are estimating the same underling quantity (sometimes called estimation error). Because of the uncertainty that is inherent in study results, physicists have understood the need for statistical methods to understand replication since at least the work of Birge (1932). This is equally true in education or the social sciences (see, e.g., Hedges and Schauer, 2019). Statistical work with replicated studies in all fields shows that even in fields with strong theory and well-developed experimental methods (like physics), the results of experiments often differ by more than would be expected solely due to chance. For example, the results of the studies estimating the universal gravitational constant depicted in Fig. 23.1 are statistically significantly different at the $p < 0.001$ level yet were judged similar enough to constitute replications and be used to establish reference values.

This does not mean that there are no standards for deciding what should be counted as replication, but rather the standard of agreement must be established by each scientific field. In an address to the National Academy of Sciences, Ralph J. Cicerone, the President of the Academy, argued that the fundamental question is

> Because variability [in findings] across studies is expected, how can we assess the acceptable degree of variability and when should we be concerned about reproducibility? (Cicerone, 2015).

Some fields have established quantitative standards for deciding how much variation of results across studies can be accepted while still assessing the studies to be replications of one another. For example, the Particle Data Group has adopted a standard that loosely corresponds to the true between-study variation that is 25% of the estimation error (see, Hedges, 2019). Similarly, the Cochrane Collaboration in medicine has established a similar standard that loosely corresponds to the true between-study variation that is 67% of the estimation error of the typical study result (see, Hedges, 2019).

23.2.5 Limits of replication and exploiting variation among studies

The authors argued above that even in the physical sciences, it is not always easy to know what methods and procedures must be kept constant in order to achieve direct replication. It is even harder in education research. Thus, direct replication is typically quite difficult to achieve in education. A recent analysis of 307 IES funded research grants found that although there were many conceptual replications, there were no instances of direct replication in IES funded research (Chhin *et al.*, 2018). This is not surprising, given the difficulties of assuring that all relevant conditions and procedures are kept constant from one study to the next.

A focus on direct replication may misdirect attention from a more important principle. Direct replication focuses on whether a particular research finding is reliable in the specific context in which that research has been conducted. But to be useful, a research finding must have a reasonably broad and known scope of applicability. The scope of applicability can only be discovered through conceptual replication that varies in certain aspects of the research context. While there is evidence that context sometimes matters profoundly for some research findings, there is also evidence that it may matter less for others. How much context matters for a particular research finding is fundamentally an empirical question.

This suggests an important role for conceptual replications and syntheses of conceptual replications in building knowledge in physics education. If conditions cannot be kept constant even within a research setting, then a more important research goal may be to try to understand the range of conditions under which a treatment will still yield positive effects. Systematic reviews of research using meta-analysis are well suited to answering such questions (see, e.g., Cooper *et al.*, 2019).

23.3 REPORTING PRACTICES AND RESOURCES TO FACILITATE REPLICATION

23.3.1 General resources

Education researchers now have significant guidance on key reporting items from intervention studies. Most relevant to education but with the broadest scope are the *Standards for Reporting on Empirical Social Science Research in AERA Publications* (AERA, 2006). Additional resources, focused on intervention studies, include the *Consolidated Standards of Reporting Trials* (CONSORT) checklist (Moher, 1998), its counterpart in the social sciences *CONSORT Social and Psychological Interventions* (Grant *et al.*, 2018), and the *Standard Protocol Items: Recommendations for Intervention Trials* (SPIRIT) checklist (Chan *et al.*, 2013). Collectively, these resources specify best practices in study reporting, with many reporting items directly supporting the fidelity of future direct replications or the contrasts tested in conceptual replications. These include reporting specifics around intervention implementation, the study context, the sample composition, administration of outcome measures, and effect size reporting. The authors note here that reporting that is consistent with these recommendations will also support study reproducibility.

As an example of one of these resources, the authors provide in Table 23.1 the CONSORT checklist of impact study reporting items. Intervention researchers in physics education might consider adopting the

Table 23.1
CONSORT checklist.

Title and abstract		
	1a	Identification as a randomized trial in the title
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)
Introduction		
Background and objectives	2a	Scientific background and explanation of rationale
	2b	Specific objectives or hypotheses
Methods		
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio
	3b	Important changes to methods after trial commencement (such as eligibility criteria) with reasons
Participants	4a	Eligibility criteria for participants
	4b	Settings and locations where the data were collected
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed
	6b	Any changes to trial outcomes after the trial commenced, with reasons
Sample size	7a	How the sample size was determined
	7b	When applicable, explanation of any interim analyses and stopping guidelines
Randomization:		
Sequence generation	8a	Method used to generate the random allocation sequence
	8b	Type of randomization; details of any restriction (such as blocking and block size)
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers) describing any steps taken to conceal the sequence until interventions were assigned
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how
	11b	If relevant, description of the similarity of interventions
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses

(Continued)

Table 23.1 ***(Continued)***
CONSORT checklist.

Title and abstract		
Results		
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned received intended treatment, and were analyzed for the primary outcome
	13b	For each group, losses and exclusions after randomization, together with reasons
Recruitment	14a	Dates defining the periods of recruitment and follow-up
	14b	Why the trial ended or was stopped
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group
Numbers analyzed	16	For each group, the number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)
Discussion		
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses
Generalizability	21	Generalisability (external validity, applicability) of the trial findings
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence
Other information		
Registration	23	Registration number and name of trial registry
Protocol	24	Where the full trial protocol can be accessed, if available
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders

CONSORT or a similar checklist to normalize impact study reporting and facilitate replication. From this checklist, the authors highlight several reporting items, discussing below only the items *most relevant to supporting direct or conceptual replication*, including *trial design, participants, interventions, outcomes, statistical methods, outcomes and estimation,* and *generalizability.* In the following sections, the authors provide brief elaboration of selected checklist items with further description later in the chapter.

Checklist Item 5: Interventions. Critical to supporting replication is clear descriptions of the educational programming delivered in both the treatment and control/comparison conditions. This could include descriptions of any secondary or supplemental interventions that may be bundled with the primary intervention. Optimally, for the treatment condition, these descriptions would include standard implementation requirements as well as allowable deviations from standard implementation (if applicable). Further, reporting the format, duration, and intensity of the primary intervention as well as for the intervention(s) in the control/comparison group is also very informative to potential replication study designers.

Checklist Item 3a: Trial design. Reporting the study design as either a randomized trial or a (non-random) quasi-experiment, along with the unit of assignment, is also important in supporting replication and facilitates optimal interpretation of effect sizes. For example, large-scale analyses of effect sizes suggest that quasi-experimental designs tend to yield larger effects than randomized designs, controlling for other study characteristics. For example, Cheung and Slavin (2016) analyzed the effect sizes from 645 studies that evaluated education interventions and found that quasi-experiments yielded effect sizes that were .07 standard deviations larger ($p = .002$) on average than those from randomized designs.

Checklist Items 4, 21: Participants/Generalizability. Information about the sample size and composition as well as the study setting will assist researchers with future replication design. Reporting sample composition entails reporting aggregate proportions of any characteristic of a study participant of interest, with particular attention to those that might be correlated with intervention outcomes. Information about study size and sample composition can also assist in the effect size interpretation. For example, one should expect effect size differences across conceptual replications that vary the sample size or composition from the original study. Cheung and Slavin (2016) observed significantly larger effects for smaller studies, on average, while Hill *et al.* (2008) found systematically larger effects for education interventions targeted at middle school students.

Checklist Item 6: Outcomes. To fully support replication, study authors should provide the full outcome measure as part of the study report (e.g., in an appendix) when possible. If a measure is proprietary, a link to the official source for that measure is desirable. When the measure itself cannot be obtained through any means, specified learning goals and/or a test blueprint for the measure provide useful information. Also important to report is the time interval between the end of the intervention and the collection of outcome data. This is relevant to replication as outcomes are often not maintained after a specified period (e.g., Atteberry and McEachin, 2021), making problematic comparisons of immediate and lagged intervention impacts (i.e., across replication studies).

Checklist Items 12, 17: Statistical methods/Outcomes and estimation. In principle, optimal statistical reporting would allow for the reproduction of results by another researcher who had access to the raw data. Statistical considerations of interest to the replication researcher include treatment of missing data and whether the analysis is "intent to treat" where all assigned units who have outcome data are included in the analysis sample, regardless of their level of treatment uptake, and participants retain their original treatment status regardless of movement between conditions. Additional helpful specifications include how nested or hierarchical data structures were addressed, whether effect sizes were based on adjusted means (and how those means were adjusted), and the variance used to standardize the effect size. More on this matter will be provided in the following section.

23.3.2 Effect size reporting

Reporting of effect sizes for intervention impacts facilitates knowledge accumulation by providing a metric of the effect that can be compared and combined across studies. Grissom and Kim (2012) indicate that the most common effect size for continuous outcomes in the social sciences is the standardized mean difference (SMD) such as Cohen's *d* (Cohen, 1977) and Hedges' *g* (Hedges, 1981). While reporting the SMD for intervention studies is now a well-established practice in education, there still exists a significant variation in how the SMD is calculated by primary researchers. In this chapter, the authors extend the general recommendation of reporting the effect sizes to recommend that the field standardize practice such that reported SMDs be standardized on the total outcome variance, which may have different sources based on the data structure. Furthermore, the authors suggest that reporting all outcome variances (within-cluster, between-cluster, total) is a prudent practice, as it provides meta-analysts with multiple options for effect size standardization.

Inaccurate effect size estimation is a key threat to statistical conclusion validity (Shadish *et al.*, 2002). A consistent approach of using the total variance to standardize the SMD (i.e., standardize the SMD denominator) maximizes the likelihood that the effect size estimate is not an overestimate or underestimate of the true effect. For studies with individual-level assignment, the total variance would be the pooled (sample size weighted) variance across treatment groups. For studies with cluster-level assignment, the total variance would be the sum of the within- and between cluster variance components estimated from a multilevel model. Details of the computation are provided in Hedges (2007).

Finally, the accumulation of knowledge about intervention effects is influenced by factors that color the interpretation of effect sizes. For example, the precision of an SMD estimate, expressed by its variance or standard error, is primarily a function of study sample size, with larger studies having smaller standard errors (i.e., more precise). In meta-analyses, effect sizes are weighted such that larger studies with more precise effect estimates have more influence on the meta-analytic average effect size. The precision of an effect size estimate also provides insight into the statistical significance of the effect and the range of values that the SMD estimate might take if the study were replicated multiple times. Other factors that are important to note while interpreting effect sizes or syntheses of effect sizes are the alignment of the outcome measure constructs with the intervention constructs, and the reliability of the outcome measure. More specifically, outcomes overaligned to treatments tend to be associated with larger effect sizes (Cheung and Slavin, 2016; and Wolf *et al.*, 2020), while outcomes with poor reliability tend to be associated with attenuated effect sizes, all else being equal (e.g., Baugh, 2002).

23.3.3 Description of intervention and comparison conditions

The authors' recommendation is for researchers to standardize how they define, label and implement intervention components. Dent and Hoyle (2015) showed that ambiguity in definitions, labels, or implementation can lead to studies being unduly included/excluded from meta-analyses, and that in small syntheses (in particular) changes in the eligibility status of just a few studies can change the substantive conclusions about intervention effectiveness. Furthermore, entities that validate education interventions, such as *Blueprints for Healthy Youth Development* (University of Colorado Boulder, 2022), require that impact study reports meet minimum levels of specificity when describing the studied interventions, before those interventions can be certified with even the lowest tier intervention rating. Similarly, a call for detailed descriptions of implementation and associated fidelity levels is echoed in the reflections of Ding (2019), who advocated comprehensive fidelity reporting toward stronger theory-building in physics education. Finally, specificity in describing comparison conditions is also crucial for interpreting the effect sizes or other impact estimates. Treatment effects must be defined by the contrast between treatment and comparison conditions. Otherwise, the effects are meaningless.

23.3.4 Transparent and comprehensive reporting of findings

Transparency is a core value of science and is essential for replication. Attention to transparency and reproducibility of research practices has gained much attention in and out of education in the last 5 years, as embodied in the *Open Science Framework* (see Foster and Deardorff, 2017). The open science movement has focused at least partly on factors that have inhibited the accumulation of knowledge. For example, publication bias, a characteristic of the research literature suggesting a systematic underrepresentation of effects that are not statistically significant (often small effects in small studies), has been well-documented (Ferguson and Heene, 2012; and Polanin *et al.*, 2016). Either through knowledge of this publication tendency or through their own experience, researchers often do not report impacts that are not statistically significant. Researchers choosing not to report non-significant effects leads to an outcome reporting bias that can lead to inflated and overly optimistic meta-analytic results (Pigott *et al.*, 2013). The pernicious effects of not being able to learn from null effects are highlighted in recent physics education studies (Conlin *et al.*, 2019).

The authors advocate for journal editorial policies that respond to this bias by not basing article-acceptance decisions on the statistical significance of observed effects. Specifically, the authors support the recommendations of van der Zee and Reich (2018) where impact studies can receive an "accepted in principle" editorial disposition after the study authors submit an acceptable literature review and methods section. Further, as a short-term strategy to promote transparency, the authors highlight the utility of study registries in education such as the *Registry of Efficacy and Effectiveness Studies* (REES: Spybrook *et al.*, 2019). Building on its predecessors in medicine, this registry requires pre-registration and eventual reporting of impacts on all planned outcomes. As the near-future will likely continue to suffer from the outcome reporting bias and publication bias, meta-analysts can also consult study registries such as these for additional effects to include in their syntheses, effects that might not otherwise be disseminated.

23.4 INSTITUTIONALIZING PRACTICES FOR MORE RAPID KNOWLEDGE ACCUMULATION

To institutionalize more comprehensive reporting practices that facilitate replication and better meta-analyses, rigorous standards will be needed. The adoption of such standards to an extent that can make measurable progress in knowledge accumulation will require community-wide buy-in. That is, multiple stakeholders must make a commitment to the standards, including researchers, journal reviewers/editors, funders, and organizations that review the trustworthiness of intervention research. In this section, the authors outline three ways to achieve this vision.

23.4.1 Funding agencies

Funding agencies can encourage more rigorous impact study designs, such as randomized experiments, by prioritizing such designs in funding decisions. However, as not all interventions can be plausibly studied with a randomized design, such a prioritization policy would have to acknowledge instances when quasi-experimental designs are the most rigorous design possible. In their funding solicitations and other guidance documents (see Common Guidelines for Education Research and Development; IES/NSF, 2013), U.S. federal funding agencies such as the Institute of Education Sciences (IES, 2022a) and the National Science Foundation (NSF, 2022) have already expressed preference for conducting randomized designs, when possible, for impact studies in education. As described above, other European agencies have expressed similar preferences (e.g., UK-based Educational Endowment Foundation/Sutton Trust, the International Initiative for Impact Evaluation, and the German National Science Foundation). The authors urge that all countries conducting physics education research consider expressing similar preferences for studies where causal inference about the effectiveness of physics education programs/practices is the primary goal, and where random assignment is plausible.

23.4.2 Preeminent journals

Editors and reviewers of the preeminent journals (e.g., *Physical Review Physics Education Research*) that publish impact studies in physics education can support higher standards by prioritizing studies with rigorous randomized designs and comprehensive reporting of findings when making article acceptance decisions. Implementing such a strategy could begin by requiring journal reviewers who review impact studies to be certified in review standards for such studies (e.g., What Works Clearinghouse Group Design Standards v. 4.1). Further, reporting elements such as those described previously in the CONSORT checklist could be incorporated into journal author guidelines in either checklist or rubric format. As indicated above, more journals that offer conditional acceptance of rigorous impact study manuscripts would likely accelerate the accumulation of impact evidence and reduce publication bias in meta-analyses. Finally, preeminent journals in physics education could stimulate and encourage synthesis work by compiling focused collections that feature meta-analyses, similar to recent efforts in engineering education (see Journal of Engineering Education, 2021).

23.4.3 Evidence clearinghouses

Evidence clearinghouses such as the U.S.-based entities *What Works Clearinghouse* (IES, 2022b), *Blueprints for Healthy Child Development* (University of Colorado, 2022), and the U.K.-based *What Works Center for Education* (Education Endowment Foundation, 2022) also play a critical role in this charge by maintaining high standards for study design and reporting and by establishing tiers of study ratings that inform decision-making by effectively distinguishing the trustworthiness of findings across studies. These study rating tiers are most influential to decision-makers when clearinghouses become the trusted resources to aid adoption decisions for programs, policies or practices. As such, an implicit incentive for researchers to use rigorous designs and to engage in comprehensive reporting of impacts is that their studies are more likely to be positively reviewed by trusted evidence clearinghouses and to subsequently facilitate better informed programmatic decisions.

23.4.4 Other incentives

In addition to receiving prioritized funding and/or publication dispositions, the authors recommend that researchers using rigorous designs and reporting findings comprehensively/transparently receive additional benefits. One incentive could be a badge system that acknowledges the efforts of researchers in these areas of research conduct. Such a badge system should convey real professional meaning to researchers, perhaps garnering extra academic credit toward promotion and tenure, and/or carry additional weight toward article- or author-level academic impact metrics. Badges could be conveyed by any and all combinations of funders, journals, and/or clearinghouses.

23.4.5 Closing comments

Knowledge creation in any scientific field requires high methodological standards to encourage valid research and transparent communication of methods/results, as well as to ensure scrutiny and replication by other scientists. The history of both the natural and social sciences suggests that the knowledge creation enterprise must rely on peer reviewed publication outlets for primary research and also programs of systematic research review to assemble, assess, and synthesize findings. Specialized outlets such as academic journals focused on publishing reviews are often part of that process. As the volume of empirical research results increases, statistical methods for summarizing research results across studies begin to play a role. Specialized centers (such as research clearinghouses) to evaluate and disseminate research findings have also become important to support systematic reviews of research and dissemination of findings.

REFERENCES

Atteberry, A. and McEachin, A., Am. Educ. Res. J. **58**(2), 239–282 (2021).

Baker, M., Nature **533**, 452–454 (2016).

Baugh, F., Educ. Psychol. Meas. **62**(2), 254–263 (2002).

Birge, R. T., Phys. Rev. **40**, 207–227 (1932).

Bollen, K. *et al.*, *Reproducibility, Replicability, and Generalization in the Social, Behavioral, and Economic Sciences. Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences* (National Science Foundation, Arlington, VA, 2015).

Bothwell, L. E. *et al.*, N. Engl. J. Med. **374**(22), 2175–2181 (2016).

Camerer, C. F. *et al.*, Science **351**, 1433–1436 (2016).

Campbell Collaboration (Campbell Collaboration, 2002). https://www.campbellcollaboration.org/

Chan, A.-W. *et al.*, Ann. Int. Med. **158**(3), 200–207 (2013a).

Cheung, A. and Slavin, R., Educ. Res. **45**(5), 283–292 (2016).

Cheung, A. *et al.*, J. Res. Sci. Teach. **54**(1), 58–81 (2017).

Chhin, C. S. *et al.*, Educ. Res. **47**(9), 594–605 (2018).

Cicerone, R. Research Reproducibility, Replicability, Reliability. President's Address to the National Academy of Sciences (April, 2015). http://www.nasonline.org/news-and-multimedia/video-gallery/152nd-annual-meeting/presidents-address.html

Cohen, J., *Statistical Power Analysis for the Behavioral Sciences* (Routledge, 1977).

Collins, H., *Changing Order: Replication and Induction in Scientific Practice* (University of Chicago Press, Chicago, 1992).

Collins, F. S. and Tabak, L. A., Nature **505**, 612–613 (2014).

Conlin, L. D. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020104 (2019).

Cooper, H. C., Rev. Educ. Res. **52**(2), 291–302 (1982).

Cooper, H. *et al.*, *The Handbook of Research Synthesis and Meta-Analysis*, 3rd ed. (The Russell Sage Foundation, New York, 2019).

Cronbach, L. J., Am. Psychol. **30**(2), 116 (1975).

Cronbach, L. J., New Dir. Program Eval. (15), 49 (1982).

Dent, A. L. and Hoyle, R. H., Metacogn. Learn. **10**, 165–179 (2015).

Ding, L., Phys. Rev. Phys. Educ. Res. **15**(2), 020101 (2019).

Docktor, J. L. and Mestre, J. P., Phys. Rev. Spec. Top. Phys. Educ. Res. **10**(2), 020119 (2014).

Edovald, T. and Nevill, C., ECNU Rev. Educ. **4**(1), 46–64 (2021).

EEF, The Education Endowment Fund (2022), see https://educationendowmentfoundation.org.uk/evaluation/about-eef-evaluation.

Ferguson, C. J. and Heene, M., Perspect. Psychol. Sci. **7**(6), 555–561 (2012).

Foster, E. D. and Deardorff, A., J. Med. Libr. Assoc.: JMLA **105**(2), 203–206 (2017).

Fraser, J. M. *et al.*, Rep. Progr. Phys. **77**(3), 032401 (2014).

Furtak, E. M. *et al.*, Rev. Educ. Res. **82**(3), 300–329 (2012).

Gardner, A. *et al.*, *A Nation at Risk: The Imperative For Educational Reform. An Open Letter to the American People. A Report to the Nation and the Secretary of Education* (National Commission on Excellence in Education, Washington, DC, 1983), ERIC Document Number ED226006.

Glass, G. V., Educ. Res. **5**(10), 3–8 (1976).

Glass, G. V. *et al.*, *Meta-Analysis in Social Research* (Sage Publications, Newbury Park, CA, 1981).

Grant, S. *et al.*, Trials **19**, 406 (2018).

Grissom, R. J. and Kim, J. J., *Effect Sizes for Research: Univariate and Multivariate Applications* (Routledge, 2012).

Hedges, L., J. Educ. Stat. **6**(2), 107–128 (1981).

Hedges, L. V., J. Educ. Behav. Stat. **32**(4), 341–370 (2007).

Hedges, L. V., J. Res. Educ. Effective. **11**(1), 1–21 (2018).

Hedges, L. V., Methodology **15**(Suppl.), 4–15 (2019).

Hedges, L. V. and Cooper, H., *The Handbook of Research Synthesis and Meta-Analysis*, edited by H. Cooper *et al.* (Russell Sage, New York, 2009).

Hedges, L. V. and Olkin, I., *Statistical Methods for Meta-Analysis* (Academic Press, New York, 1985).

Hedges, L. V. and Schauer, J., Psychol. Methods **24**, 557–570 (2019).

Hill, C. J. *et al.*, *Child Development Perspectives*, **2**: 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

Ioannidis, J. P. A., J. Am. Med. Assoc. **294**, 218–228 (2005).

IES, Institute of Education Sciences (2022a), see https://ies.ed.gov/.

IES, What Works Clearinghouse (2022b), see https://ies.ed.gov/ncee/wwc/.

IES/NSF, Guidelines for Education Research and Development (2013), see https://ies.ed.gov/pdf/CommonGuidelines.pdf.

International Initiative for Impact Evaluation. 3ie (2022). https://www.3ieimpact.org/

Jackson, G. B., Rev. Educ. Res. **50**, 438–460 (1980).

Journal of Engineering Education, Systematic reviews and meta-analyses in engineering education (2021), see https://onlinelibrary.wiley.com/doi/10.1002/jee.20415.

Kaestle, C. F., Educ. Res. **22**(1), 23–31 (1993).

Kass, E. H., *Coping with the Biomedical Literature*, edited by K. S. Warren (Praeger, New York, 1981), pp. 79–91.
Kowalski, S. and Taylor, J., *Data and Syntax for Investigating Science Education Effect Sizes: Implications for Power Analyses and Programmatic Decisions: Read Me Overview of Syntax and Data Files Associated with Taylor et al. 2018 AERA Open.Docx* (Inter-University Consortium for Political and Social Research [distributor], Ann Arbor, MI, 2019). 2019-03 23, see
Madsen, A. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **11**(1), 010115 (2015).
Makel, M. C. and Plucker, J. A., Educ. Res. **43**(6), 304–316 (2014).
McNutt, M., Science **343**, 229 (2014).
Medical Research Council, Br. Med. J. **2**, 769–782 (1948).
Moher, D., J. Am. Med. Assoc. **279**(18), 1489–1491 (1998).
Mohr, P. J. *et al.*, J. Phys. Chem. Ref. Data **45**, 1–73 (2016).
Mulrow, C. D., Ann. Intern. Med. **106**(3), 485–488 (1987).
Murnane, R. J. and Levy, F., *Teaching the new basic skills: Principles for educating children to thrive in a changing economy* (New York: The Free Press, 1996).
National Science Foundation and Institute of Education Sciences, *Companion Guidelines on Replication and Reproducibility in Education Research* (National Science Foundation and Institute of Education Sciences, Washington, DC, 2018), see https://www.nsf.gov/pubs/2019/nsf19022/nsf19022.pdf.
Olive, K. A. *et al.*, Chin. Phys. J. C **38**, 090001 (2014).
Open Science Collaborative, Science **349**, 943–951 (2016).
O'rourke, K., J. R. Soc. Med. **100**(12), 579–582 (2007).
Packer, A., *Why numbers count: Quantitative literacy for tomorrow's America*, edited by L. Steen (New York: The College Board, 1997), pp. 137–154.
Peng, R. D., Science **334**(6060), 1226–1227 (2011).
Perrin, S., Nature **507**, 423–425 (2014).
Pigott, T. D. *et al.*, Educ. Res. **42**(8), 424–432 (2013).
Polanin, J. R. *et al.*, Rev. Educ. Res. **86**(1), 207–236 (2016).
Rosenfeld, A. H., Ann. Rev. Nucl. Sci. **25**, 555–598 (1975).
Ruiz-Primo, M. A. *et al.*, Science **331**(6022), 1269–1270 (2011).
Schmidt, S., Rev. Gen. Psychol. **13**, 90–100 (2009).
Shadish, W. R. *et al.*, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Houghton Mifflin, Boston, MA, 2002).
Shavelson, R. J. and Towne, L., *Scientific Research in Education* (The National Academy Press, Washington, DC, 2002).
Shymansky, J. A. *et al.*, J. Res. Sci. Teach. **20**(5), 387–404 (1983).
Slavin, R. E. *et al.*, J. Res. Sci. Teach. **51**(7), 870–901 (2014).
Spybrook, J. *et al.*, J. Res. Educ. Effective **12**(1), 5–9 (2019).
Taasoobshirazi, G. and Carr, M., Educ. Res. Rev. **3**(2), 155–167 (2008).
Taylor, J. *et al.*, J. Res. Sci. Teach. **53**(8), 1216–1231 (2016).
Taylor, J. A. *et al.*, AERA Open **4**(3), 233285841879199 (2018).
University of Colorado Boulder, Blueprints for healthy youth development (2022), see https://www.blueprintsprograms.org/.
U.S. Congress. Education Sciences Reform Act of 2002. [Government]. U.S. Government Publishing Office (2002).
Uzunboylu, H. and Asiksoy, G., Proc. Soc. Behav. Sci. **136**(2014), 425–437 (2014).
van der Zee, T. and Reich, J., AERA Open **4**(3), 233285841878746 (2018).
Wolf, R. *et al.*, J. Res. Educ. Effective **13**(2), 428–447 (2020).
ZIB, Zentrum für internationale Bildungsvergleichsstudien (2022), see http://zib.education/en/research.html.

CHAPTER

24 QUANTITATIVE METHODS IN PER

John Stewart, John Hansen, and Lin Ding

Stewart, J., Hansen, J., and Ding, L., "Quantitative methods in PER," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 24-1–24-32.

24.1 INTRODUCTION

Physics Education Research (PER) was initially developed by practitioners traditionally trained in physics who turned their research interests to the unique problems associated with the teaching and learning of physics. As such, it is unsurprising that quantitative research is an important part of PER. Quantitative research is characterized by the collection of data which is measured numerically and the application of statistical methods to characterize that data and to understand relations in the data. This methodology differs from qualitative methods, which generally collect richer but much smaller data sets which are not amenable to statistical analysis. These two methods work to inform each other; quantitative studies suggest areas where qualitative research is needed and help determine the areas where the conclusions of qualitative research are generalizable. Qualitative research suggests areas where new quantitative studies might be productive or ways in which quantitative studies can be refined.

Quantitative methods can be separated into two broad categories: measurement and structural analysis (Ding, 2019). Measurement involves the identification of the underlying constructs important to describing an educational system. Often these constructs, such as scientific identity or Newtonian knowledge, are not directly measurable. The quantitative variables characterizing the construct are *latent variables*, or not directly measurable variables. Structural analysis develops models relating both directly measurable variables and latent variables to develop and test theories about educational phenomena. As such, measurement involves how a set of extant and latent variables are measured, while structural analysis models how these variables are related. Generally, measurement and structural analysis have been conducted independently in PER. The quality of measurement and structural models must be validated with statistical inference. Recently, a statistical method that allows the combining of measurement and structural models, which is common in broader educational research, Structural Equation Modeling, has begun to be employed in PER.

https://doi.org/10.1063/9780735425514_024

Because of the importance of quantitative methods in PER, the many excellent review articles written about the field can provide a more fine-grained overview than can be presented here. Ding and Lui (2012) provided an introductory overview of quantitative methods in PER. Ding and Beichner (2009) provided a more targeted introduction to using quantitative methods to analyze multiple-choice instruments. Ding (2019) provided a nuanced theoretical exploration of quantitative methods in PER. Docktor and Mestre (2014) provided an extensive general overview of work in PER. McDermott and Redish (1999) presented a detailed summary of early work in the field. Madsen *et al.* (2017) compiled a review of the use of research-based instruments which have been used in many PER studies. These instruments are often used in quantitative studies to infer the effects of instructional interventions on student learning outcomes; Meltzer and Thornton (2012) reviewed research on active learning methodologies. This chapter presents an overview of quantitative methods common to PER. In each section, the method will be introduced, general references are given to allow a reader to fully research the method, and then some brief examples of the use of the method in PER are presented. These examples are not and cannot be exhaustive.

24.2 RESEARCH-BASED CONCEPTUAL INSTRUMENTS

A substantial subset of quantitative work in PER involves using data collected with research-based conceptual instruments. The explosive growth of PER beginning in the 1990s can be traced to the development of research-based conceptual instruments (RBIs). The development of RBIs began with the Halloun and Hestenes' (1985b) observation that students come to physics classes with a set of stable misconceptions derived from their personal experiences and that these misconceptions generally persist after traditional instruction. This led to the assembly of a catalog of common mechanics misconceptions (Halloun and Hestenes, 1985a). This research motivated Hestenes *et al.* (1992) to develop the Force Concept Inventory (FCI) to measure a "Newtonian force concept" and to also present students with responses representing common mechanics misconceptions. Hestenes and Jackson (2010) compiled a taxonomy of misconceptions measured by the FCI. The FCI was quickly adopted at a number of institutions, which allowed Hake (1998) to gather FCI data from the introductory classes in a broad range of physics programs. These data showed that the traditional instruction was generally ineffective in promoting the learning of conceptual mechanics. The Hake study provided the evidence to support an effort, still ongoing, to bring the interactive instruction to all physics classrooms. In the 30 years since the introduction of the FCI, the evidence for the efficacy of research-based interactive instruction has become compelling (Schroeder *et al.*, 2007; Freeman *et al.*, 2014; and Von Korff *et al.*, 2016). Unfortunately, this mountain of evidence has only led to the partial adoption of these methods by the physics community (Henderson *et al.*, 2012).

The RBIs most widely used in PER studies are four measures of conceptual understanding of introductory physics: the FCI, the Force and Motion Conceptual Evaluation (FMCE) (Thornton and Sokoloff, 1998), the Conceptual Evaluation of Electricity and Magnetism (CSEM) (Maloney *et al.*, 2001), and the Brief

Electricity and Magnetism Assessment (BEMA) (Chabay and Sherwood, 1997; and Ding *et al.*, 2006), as well as two measures of students' attitudes toward different physics activities: the Colorado Learning Attitudes about Science Survey (CLASS) (Adams *et al.*, 2006) and the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) (Zwickl *et al.*, 2014). For current versions of these and many other assessments, visit PhysPort (2022).

The FCI is a 30-item multiple-choice assessment measuring a student's knowledge of Newtonian mechanics. Each item has 5 responses; the incorrect responses were developed to present the students with answers corresponding to common misconceptions. The instrument covers 1- and 2-dimensional kinematics, Newton's laws, and circular motion. The FCI contains five groups of blocked items: {5, 6}, {8, 9, 10, 11}, {15, 16}, {21, 22, 23, 24}, and {25, 26, 27}. A blocked or chained item set is a group of items all referring to each other or to a common stem. Multidimensional Item Response Theory (MIRT) studies by Stewart *et al.* (2018) show that the FCI also contains 4 groups of isomorphic items: {4, 15, 16, 28}, {5, 18}, {6, 7}, and {17, 25}. Two items are isomorphic if they can both be solved by the same physical reasoning.

The FMCE (Thornton and Sokoloff, 1998) is a 43-item multiple-choice instrument measuring a student's conceptual knowledge of one-dimensional kinematics. A revised instrument added four items measuring the understanding of energy; however, these items are rarely included in studies. The 43 items present the students with between 6 and 9 responses. Unlike the FCI, the FMCE contains items explicitly testing graphical reasoning. The FMCE uses extensive blocking of items with all items except one included in an item block. After its introduction, Thornton *et al.* (2009) provided a modified scoring rubric for the instrument, which removed some items and scored some items as groups producing a total score of 33. The FMCE contains one item that is very similar to an item in the FCI.

The CSEM is a 32-item multiple-choice assessment which broadly tests understanding of conceptual electricity and magnetism (Maloney *et al.*, 2001). Each item has 5 responses. The instrument covers electrostatics, magnetostatics, electric potential, and magnetic induction. The instrument includes 3 item blocks: {3, 4, 5}, {10, 11}, and {17, 18, 19}.

The BEMA is a 30-item multiple-choice instrument that covers conceptual electricity and magnetism (Chabay and Sherwood, 1997; and Ding *et al.*, 2006). Two-items are scored jointly to yield either 1 or 0 points. Like the CSEM, it covers electrostatics, electric potential, magnetostatics, and magnetic induction; however, it also contains 6 items covering electric circuits and 4 semi-quantitative items. The items have differing numbers of responses, with some items presenting the student with up to 10 responses. The instrument contains 7 item blocks: {1, 2, 3}, {4, 5}, {8, 9}, {14, 15, 16}, {21, 22}, {26, 27}, and {28, 29}. Despite the fact that BEMA is a broad survey of diverse electricity and magnetism topics, the items by and large cohere together to function as a measurement of a single construct (Ding, 2014a). The version available at PhysPort (2022) suggests a modified grading rubric where some items are graded as groups and some items contingently. The BEMA contains 5 items with strong similarity to items in the CSEM.

Two additional instruments that measure attitudes toward physics have been broadly applied in PER. The Colorado Learning Attitudes about Science Survey (CLASS) is a 42-item multiple-choice instrument measuring a student's scientist-like attitudes to the learning of physics and general features of physical knowledge (Adams *et al.*, 2006). The instrument uses a 5-point Likert scale. Items are scored as expert-like, non-expert-like, and neutral by comparing the student response with a panel of experts. Neutral responses were excluded from analysis. Early work on student attitudes about physics often used the Maryland Physics Expectations survey (MPEX) (Redish *et al.*, 1998). The Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) is a 31-item instrument measuring a student's attitude toward experimental physics (Zwickl *et al.*, 2014). The instrument asks the student to rate statements on both their personal beliefs and on how they believe expert physicists would answer. Again, each item is scored by comparing student responses with those of an expert panel.

Some studies have provided a comparison of the instruments. Thornton *et al.* (2009) compared pretest and post-test results of the FCI and the FMCE for a large sample drawn from two institutions and found a correlation of 0.78 between the scores of the instruments; however, the coverage of the instruments is different. Overall, 22 of the 30 FCI items would be outside the topics measured by the FMCE.

Pollock (2008) compared the CSEM and the BEMA, showing that they were fairly equivalent measures of conceptual understanding with somewhat different coverage. Xiao *et al.* (2019) provided a method for comparing BEMA and CSEM scores. Eaton *et al.* (2019b) used item response theory and classical test theory to show that the instruments had nearly equal overall difficulty.

Multidimensional item response theory constructed detailed models of all four instruments showing in detail that the instruments were quite different (Stewart *et al.*, 2018; Yang *et al.*, 2019; Zabriskie and Stewart, 2019; and Hansen and Stewart, 2021).

24.3 DESCRIPTIVE AND INFERENTIAL ANALYSES

Statistical methods can be broadly divided into descriptive and inferential analyses.

24.3.1 Descriptive analysis

A quantitative analysis generally begins with an exploration and reduction of the data, often by the calculation of a set of descriptive statistics. Data can be explored visually using scatterplots or bar charts depending on the type of data. This is an important step because problems in the data set such as outliers and data records not representative of the population may be identified and removed at this point. This may also help identify variables that have distributions that depart from the normal distribution and require special handling.

The next step is the calculation of descriptive statistics for each variable. The most common statistics are measures of central tendency: the mean, median, and mode. Most PER studies report the mean of the continuous variables. A measure of the variability of each variable should also be reported. For variables with an approximately normal distribution, the standard deviation, SD, is often reported. For variables who depart from the normal distribution, a 95% confidence interval may be more appropriate. The reporting of confidence intervals is growing in popularity in the broader education research community, as discussed in Sec. 24.8. Some studies also report the standard error of the mean, SE. The standard error is more useful in comparing means than the standard deviation and characterizes the precision with which the mean is known. If the sample is normally distributed, the central limit theorem states that $SE = SD/\sqrt{N}$ where N is the sample size. It is important to be clear when reporting either SD or SE which measure is being reported.

A researcher must also determine at this stage how to handle any records with missing entries or missing data. The most common method in PER is to delete all records with missing data. Educational data are rarely missing at random. For example, when using a pretest/post-test design, one may often be missing either the pretest or the post-test because the student was absent on the day the assessment was given. Deleting records missing either the pretest or the post-test removes students who are more likely to be absent, generally producing a data set with higher pretest and post-test scores than the class as a whole. Some biases arising from missing data are largely inevitable; however, it is important to acknowledge how the data set analyzed is related to the full sample. Other methods for handling missing data exist but are rarely applied in PER. It is possible in some cases to impute (infer) the missing values. For a discussion of multiple imputation, see Nissen *et al.* (2019).

Beyond measures of central tendency, a descriptive analysis may also include measures of association, usually the correlation between two variables. For continuous variables, the correlation coefficient r_{XY} is defined as

$$r_{XY} = \frac{E[X-\mu_X]E[Y-\mu_Y]}{\sigma_X\sigma_Y} \tag{24.1}$$

where X and Y are continuous random variables, $E[X]$ is the expectation value, μ_i is the average of variable i, and σ_i is the standard deviation. The expectation value of a random variable X is the average value of X such that $E[X] = \sum_i X_i/N$ where N is the number of observations. Correlations can be calculated between any combination of continuous and dichotomous (two level) variables. The point-biserial correlation is used for the correlation between a continuous and dichotomous variable; the phi coefficient is used for the correlation between two dichotomous variables. For a thorough discussion of correlation, see Cohen *et al.* (2003).

For a limited set of variables, the correlation matrix (the matrix formed of all pair-wise correlations) can be presented in table form. For more complex data sets, one may wish to turn to a visualization of the correlation matrix. The qgraph and corrplot packages in the R software system provide two useful visualizations of the correlation matrix.

Before proceeding with inferential analysis, one should explore the distributions of the random variables to be analyzed, so a valid inferential method can be selected. Many inferential methods assume that continuous variables are normally distributed. The assumption of normality can be investigated by examining the quantile-quantile (q-q) plot or using a quantitative test such as the Shapiro–Wilk test. Note that large data sets almost always fail a test of normality; as such, one should in general plan on testing conclusions with more robust methods (Gelman, 2021). If departures from normality are substantial, one may consider a transformation of the data.

24.3.2 Gain scores

Educational research often explores the effect of some pedagogical innovation on student learning. The amount students learn generally depends on how much they knew prior to the instruction; therefore, educational research often investigates the gain in student understanding. In PER, this is generally done by applying a pretest prior to instruction and a second application of the same instrument, a post-test, after instruction. This allows the calculation of the gain in knowledge, $Gain = Post - Pre$. In this chapter, pretest is abbreviated as *Pre* and post-test as *Post*. To allow the comparison of the application of reformed pedagogy at institutions with students with differing states of prior knowledge, Hake introduced the normalized gain g (sometimes called the Hake gain), which divided the gain by the maximum possible gain, as shown in Eq. (24.2).

$$g_i = \frac{Post_i - Pre_i}{100 - Pre_i} \tag{24.2}$$

where i indexes the student. The normalized gain has been reported both as the average of g_i or computed using class averages of pretest and post-test scores. Although extensively reported, the normalized gain has been controversial since its introduction. Bao (2006) investigated differences in gain scores arising from the two methods for calculating the gain above. Marx and Cummings (2007) suggested a modified statistic which removed the singularity generated by a perfect pretest score. Stewart and Stewart (2010) showed that the statistic was relatively immune from the suppression of gain scores due to student guessing. Coletta *et al.* (2007) showed that FCI normalized gain was significantly correlated with standardized test scores (SAT); thus, the statistic does not completely allow for the comparison of gain scores across institutions with differing student characteristics. Nissen *et al.* (2018) showed the normalized gain was biased toward students with higher pretest scores and suggested an alternate gain metric based on an effect size statistic, Cohen's *d*. This suggestion was strongly opposed by Coletta and Steinert (2020), who showed that the dependence of normalized gain on the pretest score was fully explained by differences in the Lawson Classroom Test of Scientific Reasoning. Note that the reporting of normalized gain is unique to PER and can make PER studies difficult to interpret by other disciplines; as such, it is important to provide a complete report of statistics beyond the normalized gain including pretest scores, post-test scores, sample sizes, and standard deviations.

24.3.3 Inferential statistics

Inferential analysis allows the researcher to test statistical assumptions about the data. For example, one can test how likely it is that observed differences in mean scores or observed measures of association happened by chance. To perform an inferential analysis to test a hypothesis, first one states a null hypothesis H_0 and a mutually exclusive alternate hypothesis H_1. For example, if one wishes to test if two sets of post-test scores are statistically different, the null hypothesis would assert the means of the two scores were equal, H_0: $\mu_1 = \mu_2$ and the alternate hypothesis would then have the means different, H_1: $\mu_1 = \mu_2$. To test a hypothesis, one assumes that the null hypothesis is true (and makes some assumption about the distribution of the data) and calculates a test statistic with a known probability distribution. The distribution is then used to compute the probability p that a value equal to or larger than the test statistic occurred by chance. A significance threshold, α, is then selected. If $p < \alpha$, then the null hypothesis is rejected and the alternate hypothesis is accepted. For most PER studies, the significance threshold is chosen as $\alpha = 0.05$.

It is important to account for all correlations within the data set when performing inferential analysis. Care should be taken when examining differences where multiple measurements are performed on the same individual (a within group or repeated measures design) as opposed to measurements made on different groups of individuals (a between groups design). Many statistical methods have been developed to account for the additional correlations of repeated measures designs, including ANOVA (analysis of variance) and linear mixed models (West *et al.*, 2015).

24.3.4 Bootstrapping

As an alternative to hypothesis testing or other methods with a distributional assumption such as normality, one can also use a computer-based method such as bootstrapping (Efron, 1992; and Davison and Hinkley, 1997). Bootstrapping develops the sampling distribution by resampling the data set with replacement. This generates as many replications of the original distribution as desired by randomly deleting or duplicating records. Any test statistic desired can be calculated for each distribution producing the distribution of the statistic. For example, the mean of each distribution could be calculated for each distribution; if 1000 bootstrap replications were used, this would generate 1000 estimates of the mean. If the distribution of means was symmetric, the uncertainty in the mean could be characterized by calculating the standard deviation of the distribution of mean values. Bootstrapping is often used for statistics whose distribution departs from the normal distribution; in these cases, it is more appropriate to report the 95% confidence interval. Two samples are significantly different if their confidence intervals do not overlap. The "boot" package in the R software system provides support for bootstrapping and the estimation of confidence intervals (Canty and Ripley, 2017). Bootstrapping is insensitive to assumptions about the distribution of the sample because it uses the actual distribution of the sample.

24.3.5 Effect sizes

The central limit theorem guarantees that with a sufficient sample size, any difference in the sample means between two samples will be a statistically significant difference; however, the difference may not be of practical importance. To provide a measure of the functional size of either the difference in two random variables or the degree of association in two random variables, Cohen (1977) introduced the “effect size.” Effect sizes characterize differences in means or degrees of association as small effects, medium effects, and large effects. Differences or degrees of association that are statistically significant but less than a small effect are practically negligible. The most common effect size measure for the difference in the means of the two samples is Cohen’s *d, which* is defined as the difference in the means divided by the pooled standard deviation. Cohen’s d measures the difference in mean in standard deviation units. The effect size criteria for d are that 0.2 is a small effect, 0.5 is a medium effect, and 0.8 is a large effect. The correlation coefficient r measures the degree of association of two variables; the effect size criteria for r is that $r = 0.1$ is a small effect, 0.3 is a medium effect, and 0.5 is a large effect. In general, any report of a significant difference in two samples should be accompanied by the effect size of the difference. For a readable discussion of effect sizes, see Ellis (2010).

Effect sizes are guidelines which Cohen (1977) developed from qualitative considerations such as that a medium effect should be visible to the naked eye. These criteria were developed by considering what values of the effect size statistics would be seen in studies in the social sciences. When these statistics are used for other purposes, such as to characterize gain scores, the criteria should be re-examined.

24.3.6 Type I error–Bonferroni correction

A statistical test of a hypothesis is susceptible to two broad types of errors: Type I error and Type II error. A Type I error is made when the null hypothesis is rejected when it is actually true; this kind of error is a false positive where random fluctuations in the data are interpreted as a real effect. If one is using the $\alpha = 0.05$ significance threshold, then a Type I error should be made in one of 20 statistical tests performed. As such, studies using many statistical tests should correct for the inflation of the Type I error rate inherent in performing multiple statistical tests. Many methods exist to correct for the inflation of Type I error; the most commonly used in PER is the Bonferroni correction, which adjusts the significance threshold based on the number of statistical tests performed. The modified threshold is the original threshold divided by the number of tests performed, k, $\alpha' = \alpha/k$. With this modification, only tests with $p < \alpha'$ are considered statistically significant. We note that other correction methods exist, and the Bonferroni correction is one of the more aggressive; fewer p values will be found significant than with other corrections. For example, the Hochberg correction examines the p values sequentially and consistently increases the α correction as more statistical tests are performed (Andrade, 2019).

24.3.7 Type II error–statistical power

A Type II error occurs when the null hypothesis is accepted when it is actually false; this error produces a false negative. One common source of Type II errors is insufficient sample size, which leads to a lack of statistical power. Statistical power represents the probability that a statistical test will detect an effect that actually exists. Statistical power generally depends on the significance threshold, the sample size, and the size of the effect one wishes to measure. Before concluding that an effect is not statistically significant, one should perform a power analysis to determine whether the sample was of sufficient size to allow the reliable detection of the effect. Such analyses are fairly rare in PER; Stewart *et al.* (2021a) provide an example of a power analysis in a study examining underrepresented groups in physics classes where the sample had insufficient statistical power to detect small effect sizes.

24.3.8 Beyond significance testing

The correct use and interpretation of statistical methods is central to the effectiveness of PER. As such, the PER community should be up to date with the latest advancements and changes in statistical research. Recently, the topic of hypothesis testing and reporting significance determined by p values has been called into question, to the point that some journals discourage the use of null hypothesis significance testing (NHST). One journal has even banned the use of p values in its publications (Greenland *et al.*, 2016). One of the biggest arguments for the elimination of p-value reporting is that it is so poorly understood and often incorrectly interpreted that many faulty conclusions are drawn from valid statistical work (Calin-Jageman and Cumming, 2019). Greenland *et al.* (2016) summarizes many of the misinterpretations that plague p-values, as well as misinterpretations of confidence intervals and power testing. Other issues arise with the use of p-values, such as dichotomous thinking that something is significant or not significant based on a p-value being above or below a fairly arbitrary value of 0.05. This dichotomous thinking has led to a misrepresentation of statistical analyses, where only studies that find significance are reported while those that do not find significant results are not reported (Calin-Jageman and Cumming, 2019). An effect that is found to be significant in several studies may in fact be insignificant in many other studies, but because the insignificant findings were not reported, the public receives a skewed or misleading interpretation of findings. Cumming (2014) sets forth a program for nearly eliminating the reporting of p-values and significance testing, suggesting that studies focus on effect sizes and estimation. Other studies (Kubsch *et al.*, 2021) suggest the use of a Bayesian approach that focuses on posterior distributions as opposed to the frequentist NHST approach. These changes in the use of statistics for research purposes are not widespread in PER.

24.4 CLASSICAL TEST THEORY

The origins of Classical Test Theory (CTT) can be traced back to the early 20th century; CTT was developed to systematically evaluate psychological tests. Crocker and Algina (1986) provided a classic introduction to CTT. The most basic principle of classical test theory is the idea that an observed score on a test using a given participant is the sum of that participant's true score and their error score:

$$O = T + E, \tag{24.3}$$

where O is the participant's observed score, T is their true score, and E is the error score. The participant's true score is defined as the score the participant would receive if the test perfectly measured the psychological constructs or skills that it purports to measure. In other words, it is the true ability of the participant in the domain that the test measures. The error score is defined as the sum total of the errors made on the test by the participant. Some errors are positive (the participant correctly answers an item they generally would not have) while others are negative (the participant incorrectly answers an item they generally would not have). These errors result from the writing and structure of the test or the environment in which the test is administered. Generally, the goal of CTT is to minimize the error scores of participants on tests so as to lessen the gap between observed and true scores. This is done by examining two key characteristics of a test: its validity and its reliability.

24.4.1 Test validity

The first of the two characteristics examined by CTT is validity. A test is considered valid if the knowledge and skills a test claims to measure are relevant to the domain of the test, or more simply, whether or not a test actually measures what it purports to measure. There are three different types of validity; generally, the purpose of a test determines the type of validation used in its creation.

The first is content validity. Content validity is judged in three parts: relevance, balance, and specificity. The instrument's content relevance, sometimes referred to as "face validity" can be determined by having test takers or content experts examine the instrument and answer the question, "Are the questions of the instrument within the domain of the topic?" Content balance reflects the coverage of the subject matter and strives to answer the question, "Does a test cover enough or too much of the aspects of a specific topic?" Content specificity refers to whether a test measures content specific concepts, or concepts that a content expert would know and a content novice would not. A test that has good content specificity would have characteristics such that a subject matter expert should be near perfect in their score, while a participant who is completely naïve to the subject matter should get near chance scores. Content validity historically has no statistical measure and is determined by the consensus of expert opinion.

The other types of validity are criterion validity and construct validity. Criterion validity looks for evidence that performance on an instrument can be used to make inferences about performance in a different domain. For example, one could measure the criterion validity of an instrument that tested concepts of electric circuits by analyzing whether performance on the instrument predicted or concurred with performance in designing and carrying out an electric circuits lab experiment. Construct validity looks for evidence that the instrument measures abstractions such as intelligence, creativity, and problem-solving ability. Construct validity is often established by examining correlations between the construct and other established measurements to determine whether the pattern of correlations matches theoretical expectations.

In PER studies that examine the effectiveness of physics conceptual inventories, content validity is generally addressed. Some studies have addressed criterion validity as well, such as convergent and divergent validity evidence. Construct validity is often discussed in the context of Rasch analysis (Planinic *et al.*, 2010; and Ding, 2014a).

24.4.2 Test reliability

The other characteristic measured in CTT is the test reliability. Test reliability is generally determined in two parts: by measuring its consistency and discriminatory power.

Consistency is a measure of how replicable a test's results are if administered to the same population again under similar conditions. There are four methods to determine the consistency of a test. The alternate or parallel form method administers a second test that is exactly parallel to the first (test items cover the exact same concepts with parallel but not exactly similar questions). The scores on one test should be approximately the same as those on the other. The test-retest method administers the same test to the same individuals twice. The split-half method splits one test into two similar or parallel sub-parts. The score on one half of the test should be similar to the score on the other half of the test, indicating consistency. Item covariance methods are similar to the split-half method, except the test is split into more than two parts, and consistency is determined between the smaller sub-parts of the test and the rest of the test. Cronbach's alpha is an item covariance method that splits the test into the smallest possible sub-parts, i.e., the individual items. Cronbach's alpha measures the internal consistency between each item and the entire test and has the form

$$\alpha_C = \frac{k}{k-1}\left(1 - \frac{\Sigma\sigma_{xi}^2}{\sigma_x^2}\right) \tag{24.4}$$

where k is the number of items on the instrument, σ_{xi} is the variance of item i and σ_x is the total test variance. Kuder-Richardson 20 (KR-20) is equivalent to Cronbach's alpha for dichotomously scored items and has been used in the validation of physics conceptual inventories (Ding *et al.*, 2006). KR-20 has the same form of Cronbach's alpha except the item variance σ_{xi} is replaced with the dichotomously scored item variance $\sqrt{p(1-p)}$ where p is the average score of the item. KR-20 gives a good approximation of the test reliability and can be used to calculate the standard error of measurement (SEM) as $SEM = \sigma_x\sqrt{1-R}$ where σ_x is the standard deviation of test scores, and R is the test reliability. Using SEM, one can calculate the score confidence interval (CI). For example, the 95% CI is $Score \pm (1.96 \cdot SEM)$.

The other measure of test reliability is the test's discriminatory power, or how well it discriminates between individuals with high-levels of ability and individuals with low-levels of ability in the test's domain. Single items on a test can be measured for discriminatory power. The most common measures are the item difficulty index and the item discrimination index. The item difficulty is the average score on the item; un-intuitively, items with a higher difficulty index are answered correctly by more students. The item discrimination D is calculated by comparing the percentage of higher performing students who answer an item correctly the fraction of low performing students. This is shown in Eq. (24.5).

$$D = P_u - P_l, \tag{24.5}$$

where P_u is the proportion of participants with the top X% of total scores who answer the item correctly and P_l is the proportion of participants in the bottom X% answering correctly. Various values of X are used where $X = 25\%$, $X = 50\%$, and $X = 27\%$. Beyond item-level discrimination, the overall discrimination of a test can be calculated through Ferguson's delta. Ferguson's delta is a measure of the extent to which test scores spread across the entire possible range of values, thereby separating high-achieving and lower-achieving students.

Classical Test Theory has been applied by Ding *et al.* (2006) to evaluate item functioning in the BEMA. Traxler *et al.* (2018) applied CTT as part of a study of item fairness in the FCI finding multiple items unfair to women and a few unfair to men. An item is "fair" for some groups of students if students within the group and outside of the group with the same general facility with the material score equally on the item. Henderson *et al.* (2018) applied the same techniques to the FMCE and the CSEM finding few unfair items in these instruments. Eaton *et al.* (2019b) used CTT to compare the CSEM and the BEMA. Test reliability has been reported for the FCI (Lasry *et al.*, 2011) and the BEMA (Ding, 2014a) with both having high reliability.

24.5 ITEM RESPONSE THEORY

Item Response Theory (IRT) represents a broad collection of statistical models of the probability of selecting a response on a multiple-choice instrument. For an overview of IRT, see van der Linden (2016).

A wide variety of IRT models have been applied within PER. The most common model is the 2-parameter logistic model (2PL). The 2PL model predicts the probability that the correct response is selected; each multiple-choice item is dichotomously scored as either correct (1) or incorrect (0). The 2PL model assumes that each item, j, is described by a discrimination, a_j, and a difficulty, b_j. The probability, π_{ij}, that student i answers item j correctly is given by

$$\pi_{ij} = \frac{exp\left[a_j\left(\theta_i - b_j\right)\right]}{1 + exp\left[a_j\left(\theta_i - b_j\right)\right]} \tag{24.6}$$

where θ_i is a latent variable estimated for each student i measuring his or her general facility with the material tested by the instrument. In IRT, θ_i is generally called the "ability." Despite the name, the ability is specific to the construct being measured. This should not be confused with some universally general or amorphous abilities applicable to any domain or content area. The parameter θ is probably better conceptualized as a parameter related to the student's general facility with answering the items correctly on the instrument under consideration. Parameters a and b are named analogously to quantities in classical test theory and perform the same general function; however, the mathematical values of the difficulty and discrimination will not be the same for IRT and CTT. The 2PL model is of a similar form to the function used in logistic regression (Sec. 24.8.2). The expression can be converted into a linear relation by forming the odds, $odds_{ij} = \pi_{ij}/(1-\pi_{ij})$ and then taking the logarithm, as shown in Eq. (24.7)

$$ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = a_j\left(\theta_i - b_j\right) = a_j\theta_i - a_jb_j, \tag{24.7}$$

which shows $-a_jb_j$ as the intercept and a_j as the slope of the log-odds with respect to θ.

IRT has been applied in a diverse set of studies in PER. Wang and Bao (2010) applied the 3-parameter logistic (3PL) model to a large FCI pretest sample drawn for a calculus-based university mechanics course. The 3PL model generalizes the 2PL model by adding an additional parameter to account for random guessing. Han *et al.* (2015) also used the 3PL model to split the FCI into two half tests. The FCI items were generally well performing in this analysis, with difficulty parameters in the range of good item function and all discrimination parameters positive. The 3PL model was also used as part of an adaptive testing system to apply the FCI (Yasuda *et al.*, 2022). Planinic *et al.* (2010) analyzed a large sample of FCI post-test data using the 1-parameter logistic (1PL) model and found that the item difficulties were generally commensurate to the overall item average. The 1PL model is a simplification of the 2PL model, which constrains the discrimination to one ($a = 1$).

IRT has also been used to investigate item fairness. Osborn Popp *et al.* (2011) applied the 1PL model to a sample of 4775 high school students FCI scores and found a number of FCI items with different difficulty for men and women. Traxler *et al.* (2018) applied IRT to three samples of FCI pretest and post-test scores, finding five items substantially unfair to women. Additional analysis using Differential Item Functioning (DIF) identified 8 items unfair to women and 2 unfair to men.

Much less IRT work has been done with other RBIs. Henderson *et al.* (2018) replicated the analysis of Traxler *et al.* (2018) for the FMCE and the CSEM and found few items in these instruments substantially unfair to either men or women.

Unidimensional IRT estimates a single ability parameter θ_i for each student; Multidimensional IRT (MIRT) estimates K ability parameters for each student using the generalization on the 2PL model shown in Eq. (24.8),

$$\pi_{ij} = \frac{exp\left[\boldsymbol{a}_j \cdot \boldsymbol{\theta}_i + d_j\right]}{1 + exp\left[\boldsymbol{a}_j \cdot \boldsymbol{\theta}_i + d_j\right]} \tag{24.8}$$

where $\boldsymbol{a}_j$ is a K component discrimination vector and $\boldsymbol{\theta}_i$ a K component ability vector; $\boldsymbol{a}_j \cdot \boldsymbol{\theta}_i$ is the dot product of the vectors. The quantity d_j replaces $-a_jb_j$ in the 2PL model; easier items have larger d_j.

MIRT with an unconstrained discrimination matrix is very similar to the factor analysis described in Sec. 24.7. The K dimensions represent factors and the discrimination matrix $\boldsymbol{a}_j$ represents the loadings of those factors on each item. In general, in PER, unconstrained MIRT has been used to perform factor analyses on RBIs. MIRT applies maximum likelihood estimation techniques (Sec. 24.3) which provide the researcher a wealth of additional model selection statistics beyond traditional factor analysis methods. MIRT has been used to extract factor structures for the FCI (Scott and Schumayer, 2015; Stewart *et al.*, 2018; and Eaton and Willoughby, 2020), the FMCE (Yang *et al.*, 2019), the CSEM (Zabriskie and Stewart, 2019), and the BEMA (Hansen and Stewart, 2021). Constrained MIRT, which constrains the discrimination matrix to a theoretical model, has been used to provide detailed models of the skills tested by the RBIs (Stewart *et al.*, 2018; Yang *et al.*, 2019; Zabriskie and Stewart, 2019; and Hansen and Stewart, 2021).

Both Eqs. (24.6) and (24.8) predict the probability distribution of dichotomously scored items. Recently, IRT models which predict the probability of selecting each response to a multiple-choice instrument have begun to be applied to conceptual inventory data. The Nominal Response Model (NRM) (Bock, 1972) in Eq. (24.9) is an example of one such model

$$\pi_{ijk} = \frac{exp\left[\alpha_{jk} \cdot \left(\boldsymbol{a}_j \cdot \boldsymbol{\theta}_i\right) + d_{jk}\right]}{\sum_{m=1}^{R} exp\left[\alpha_{jm} \cdot \left(\boldsymbol{a}_j \cdot \boldsymbol{\theta}_i\right) + d_{jm}\right]} \tag{24.9}$$

where π_{ijk} is the probability that student i selects response k to item j, $\boldsymbol{a}_j$ is a K component discrimination vector for item j, $\boldsymbol{\theta}_i$ is a K component ability vector for student i, α_{jk} is the overall discrimination of response k to item j, d_{jk} is a parameter related to the overall probability that response k of item j is selected, and R is the number of responses. Note, α_{jk} multiplies the dot product $\boldsymbol{a}_j \cdot \boldsymbol{\theta}_i$ and thus is sensitive only to the combination of θ dimensions defined by $\boldsymbol{a}_j$ not to the individual θ dimensions. Equation (24.9) generalizes Eq. (24.8) by normalizing the probability of selecting some response to one by dividing by the sum of the Eq. (24.8) probabilities. Smith and Bendjilali (2022) showed that the discrimination parameters in the NRM can be used to rank incorrect responses to the degree they are related to correct underlying knowledge.

Eaton *et al.* (2019c) applied a combination of the unidimensional 2PL and the nominal response model (2PLNRM) to construct a partial credit model for the FCI. Smith *et al.* (2020) also applied the 2PLNRM to the FMCE to provide a ranking of the responses. Stewart *et al.* (2021b) applied NRM and cluster analysis (Sec. 24.10) to three large FCI samples.

The 1PL model is sometimes called the Rasch model and should not be confused with Rasch analysis. Rasch analysis, like IRT, places both students and items on the same scale and can be used to determine whether the item distribution is similar to the person distribution. Rasch analysis differs fundamentally from IRT in that the former is a theory-laden approach to quantitative measurement and is operated by fitting empirical data to a model, whereas the latter is centered around data and seeks to fit different models to data. This nuanced difference underscores the higher standards placed on Rasch theory, whose results can uniquely meet the requirements for objective measurement. Rasch theory offers a unique measurement framework that can afford rich information to help establish validity evidence.

Rooted in falsificationism, Rasch theory in principle represents a scientific approach where anomalies, in the form of misfit statistics, are explicitly sought to refute an a priori model. Ding (2014a) applied Rasch analysis to investigate the BEMA. For a discussion of Rasch analysis, see the chapter by Ding (2022) in this collection.

24.6 FACTOR ANALYSIS

Factor analysis is a technique which attempts to explain the variance in a set of observed variables using a smaller set of unobserved and generally unobservable variables, latent variables. Factor analysis can be applied as either an exploratory or a confirmatory technique. In Exploratory Factor Analysis (EFA), no prior factor structure is assumed; the factor structure is deduced from the data by finding the factor structure that balances model fit with parsimony. A model is more parsimonious if it uses fewer variables. In Confirmatory Factor Analysis (CFA), the researcher proposes a factor structure and evaluates how well that structure fits the observed data. CFA is often accomplished using Structural Equation Modeling (Sec. 24.9).

An EFA proposes a set of linear relations between a set of N observed variables y_{ji} representing the measurement of y_j for participant i. The variation observed in y is explained by a set of K latent traits, x_{ik}, where x_{ik} is the latent trait associated with factor k of participant i. The latent traits are related to the observed data by a set of factor loadings f_{jk}. The resulting set of equations is shown in Eq. (24.10).

$$\begin{aligned} y_{1i} &= f_{11}x_{1i} + f_{12}x_{2i} + f_{13}x_{3i} + u_{1i} \\ y_{2i} &= f_{21}x_{1i} + f_{22}x_{2i} + f_{23}x_{3i} + u_{2i} \\ &\dots \\ y_{ni} &= f_{n1}x_{1i} + f_{n2}x_{2i} + f_{n3}x_{3i} + u_{3i} \end{aligned} \tag{24.10}$$

where u_{1i} is the residual error for student i in y_1.

Table 24.1 shows the results of an EFA of the first 15 items of the FCI. This table has not been previously published and is used as an example. The first factor explained 15% of the variance in the item scores; the second factor 8%. The number of factors was selected by examining the scree plot. Scree plot plots the number of factors (or later clusters) on the horizontal axis and a measure of model fit such as the total variance explained or the additional variance explained on the vertical axis. As has been observed in most EFA work on RBIs, some factors mix different concepts in mechanics, with Factor 1 having strong loadings on Newton's 3rd law (items 4 and 15) and kinematics (items 11 and 13).

One of the most important decisions in a factor analysis is the selection of the number of factors. The factor analysis method used found the factors as eigenvectors of the correlation matrix. The eigenvalue associated with the eigenvector is related to the variance explained. The scree plot, Fig. 24.1, plots the eigenvalue against the number of factors. One method for selecting the optimal number of factors to

Table 24.1

Factor loadings for a 2-factor model of the first 15 items on the FCI.

	Factor 1	Factor 2
FCI 1		
FCI 2	0.30	
FCI 3	0.36	
FCI 4	0.50	
FCI 5		
FCI 6		0.48
FCI 7		0.42
FCI 8	0.43	
FCI 9	0.52	
FCI 10	0.37	
FCI 11	0.52	
FCI 12		0.31
FCI 13	0.56	0.31
FCI 14		0.36
FCI 15	0.65	

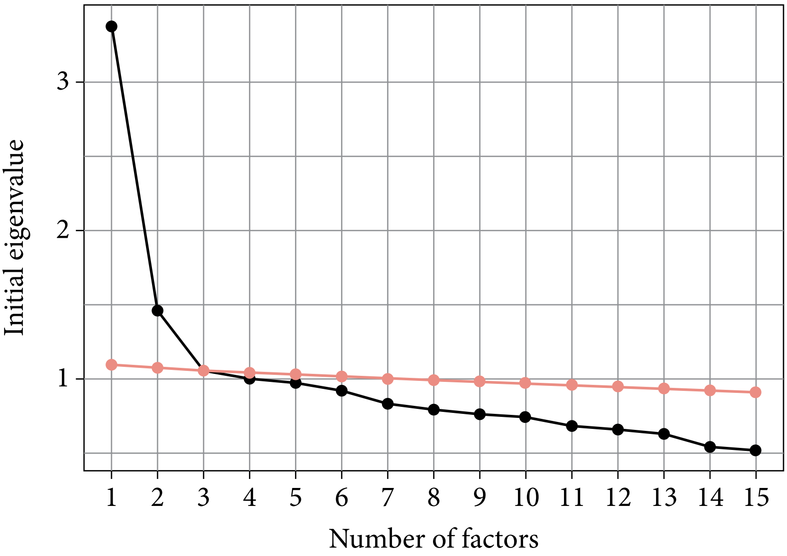


FIG. 24.1

Scree plot of the first 15 items on the FCI (black line). The red line presents the parallel analysis for 15 items.

describe the data is to look for an “elbow” or “knee” in the scree plot, the point where the plot is maximally curved. Figure 24.1 shows that the knee occurs at 2 or 3 factors. The trace of the correlation matrix is the number of items (15) because all diagonal entries are one. The sum of eigenvalues of a matrix is equal to the trace of the matrix; therefore, eigenvalues less than one represent factors which explain less variance than a single item. One goal of EFA is to explain the observed data with a smaller number of factors; therefore, factors that don’t explain as much variance as single items should be discarded. This suggests a 2-factor model; the third eigenvalue is barely above one. The red line in Fig. 24.1 presents a parallel analysis, the result of factoring a random data set of the same size as the analysis data set. The selected number of factors should be above the parallel analysis line, suggesting a 2-factor solution.

Principal Component Analysis (PCA) also seeks to explain the variance in a set of items with a more parsimonious set of components. PCA does not model these components as latent variables but is strictly a dimensional reduction technique. PCA finds the linear combination of the observed variables which explains the most variance as the first component. The variance of this component is removed, then the second component is the linear combination explaining the most remaining variance. These components are explicitly required to be uncorrelated. As above, this can be accomplished by finding the eigenvectors of the correlation matrix. As such, while philosophically different, EFA and PCA often yield very similar results. Note that the factors extracted by EFA can be correlated; those extracted by PCA are required not to be.

The results of either EFA or PCA form a set of coordinate vectors in the K-dimensional space defined by the K factors. This coordinate system can be arbitrarily rotated to be easier to interpret. Many factor rotations exist; Table I uses varimax rotations, which seeks factor loadings with a few large values and as many zeros as possible and leads to orthogonal factors. In EFA, other rotations allow factors to be correlated (this is theoretically reasonable in many cases). The goal of rotation is to find the simplest structure of the correlation matrix that gives easily interpretable results and retains all the pertinent correlations (Sass and Schmitt, 2010). The simplest structure possible is one where each item loads only on one factor, and there is no interfactor correlation. Rotation methods should be carefully chosen as many researchers oversimplify the structure through the rotation and lose valuable interfactor correlations. For a much more in-depth review of rotations in EFA and how to select a method, see Sass and Schmitt (2010).

Recently, Multidimensional (MIRT) IRT has been used to extract factor structures of many RBIs. Scott and Schumayer (2015) showed that traditional EFA and MIRT yielded similar but not identical results.

Many studies in PER have applied EFA to the RBIs introduced in Sec. 3; in general, the theoretical factor structure proposed by the authors of the instruments has not been recovered. For the FCI, this led to one of the most famous early controversies in the field. Hestenes *et al.* (1992) suggested that the FCI measured 6 dimensions of a Newtonian Force Concept; Huffman and Heller (1995) performed PCA on the FCI and showed that only two factors were identified. This led to a lively back and forth discussion with the FCI authors asserting that the instrument was not designed to have a factor structure (Heller and Huffman, 1995; and Hestenes and Halloun, 1995). Since then, multiple studies have identified 5-factor (Scott *et al.*, 2012), 6-factor (Semak *et al.*, 2017), and 9-factor solutions (Stewart *et al.*, 2018) as optimal for the FCI post-test. None of these solutions conformed to the original published structures and most mixed items made little theoretical sense together. Stewart *et al.* (2018) showed that much of the factor structure was accounted for by item blocking and a small number of isomorphic items in the instrument.

Less work has explored other RBIs. Ramlo (2008) performed EFA on a small FMCE data set finding 3-factors as optimal for the post-test. Yang *et al.* (2019) performed a factor analysis of the FMCE using MIRT; a 5-factor model was optimal for the majority of the fit statistics. This model was closely related to the blocked structure of the instrument. Maloney *et al.* (2001) used PCA to analyze the CSEM with the initial publication of the instrument yielding an 11-factor model; the model explained too little variance and was not reported. Zabriskie and Stewart (2019) used MIRT to perform EFA on CSEM samples from two institutions finding 8 and 9-factor models as optimal. Eaton *et al.* (2019a) compared the factor structures of the CSEM and the BEMA. Douglas *et al.* (2014) performed EFA on a large sample of CLASS data and found that only three factors were extracted, not the eight factors reported by the authors of the instrument (Adams *et al.*, 2006). The 3-factor model has since been confirmed in another study (Christman *et al.*, 2020).

None of these studies extracted a factor structure consistent with a published theoretical structure of the instrument and most found factors influenced by item blocking. As such, the RBIs do not seem to have a well defined subscale (factor) structure. This may limit both their practical use in the classroom and their usefulness in studies because they give one coarse-grained measure of conceptual understanding but provide little information about how that understanding varies across subtopics. For example, while one can measure an overall Newtonian force concept with the FCI, one cannot measure how components of that concept such as 1-dimensional kinematics contribute to that force concept because many items in the instrument require reasoning from multiple subtopics for their solution.

24.7 REGRESSION

Various forms of regression modeling have been extensively used in PER. Each form attempts to model the variation of a dependent variable using a linear combination of independent variables. When the dependent variables are continuous, linear regression is used. When the dependent variable is dichotomous, logistic regression is appropriate. As such, IRT is a special case of logistic regression. For a complete review of the regression analysis, see Cohen *et al.* (2003).

24.7.1 Linear regression

Linear regression models the variation of a continuous dependent variable using a linear combination of independent variables. The independent variables can be continuous, dichotomous, or categorical. An example of a linear regression model predicting the post-test score (*Post*) with the pretest score (*Pre*) and ACT mathematics percentile scores (*ACTM* %) is shown in Eq. (24.11)

$$Post = \beta_0 + \beta_1 \cdot Pre + \beta_2 \cdot ACTM\% + \epsilon \tag{24.11}$$

where β_0 is the intercept, β_1 and β_2 are the slopes, and ϵ is the residual error.

Linear regression finds an exact minimization of the sum of square errors ϵ^2. A regression analysis will also in general report R^2, the fraction of the variance explained by the regression model, the probability the model happened by chance based on the F statistic, as well as the probability each regression coefficient resulted from a random fluctuation. To calculate these probabilities, some statistical assumptions must be met: (1) the dependent variable is randomly sampled for each value of the independent variable, (2) the variance of the residual error ϵ is independent of the values of the dependent variables (homoscedasticity), (3) each observation is independent of other observations, and (4) for each value of the dependent variables, the independent variable is normally distributed. Educational data often fail to meet these assumptions. The conclusions of the regression analysis should then be checked with robust regression methods less sensitive to these assumptions. Many robust methods exist, which are less sensitive to the normality of the error distribution or the existence of outliers (Yu and Yao, 2017). If the conclusions are not supported, the research may either report the robust results or attempt a transformation of the data such that the transformed data meet the assumptions.

24.7.2 Logistic regression

Linear regression is by far the most commonly used regression method in PER, where often the goal is to understand changes in a continuous variable such as class test average or conceptual post-test score. Some studies have investigated factors affecting a dichotomous variable, such as passing a physics class or continuing in a physics major. While a dichotomous variable is numerical, normal linear regression is inappropriate to model the relationship of the dependent dichotomous variable with a set of independent dichotomous, categorical, or continuous variables. For a dichotomous variable Y, we would like to model how the probability distribution of selecting the high level of the variable $P\,(Y = 1)$ depends on a set of independent variables. To do this, an analysis very much like IRT is carried out, where the log-odds is predicted by a linear function in the dependent variables. An example is shown in Eq. (24.12), which predicts the log-odds using the pretest score and the ACT mathematics percentile

$$ln\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 \cdot Pre + \beta_2 \cdot ACTM\% \tag{24.12}$$

where β_0 is the intercept, β_1 and β_2 are the slopes. The log-odds is predicted because it has an unlimited range from $-\infty$ to $+\infty$ greatly simplifying its estimation; the odds are restricted to the range from 0 to $+\infty$.

Logistic models are a bit harder to interpret than linear models. To test whether the model is significant, one first fits a null model containing only the intercept. A chi-squared test comparing the null model and the test model will determine whether the test model is a significant improvement over the null model. To allow a more intuitive interpretation of the regression coefficients, one exponentiates Eq. (24.12) to form

$$\frac{P(Y=1)}{1-P(Y=1)} = e_0^{\beta} \cdot e^{\beta_1 \cdot Pre} \cdot e^{\beta_2 \cdot ACTM\%} \tag{24.13}$$

The ratio $\frac{P(Y=1)}{1-P(Y=1)}$ is the ratio of the probability that the event $Y = 1$ happens to the probability that the event $Y = 0$ happens is called the odds. The term e^{β_0} is the base odds, and the odds of $Y = 1$ if all the independent variables are zero. The terms $e^{\beta_1 \cdot Pre}$ and $e^{\beta_2 \cdot ACTM\%}$ are called odds ratios; they multiply the base odds. If $e^{\beta_1 \cdot Pre} = 2$ and $e^{\beta_2 \cdot ACTM\%} = 2$, the odds of $Y = 1$ are multiplied by four. Logistic regression assumes that the individual observations are independent, not co-linear and that there is a linear relationship between the independent variables and the log-odds.

Logistic regression has been used in PER to study a diverse set of topics. Logistic regression has been used to explore factors influencing the persistence of LGBT+ physicists (Barthelemy *et al.*, 2022), the factors affecting the motivations and considerations of leaving physics majors (Barthelemy and Knaub, 2020), the factors influencing women to pursue careers in physics (Hazari *et al.*, 2013), to understand the gender gap in conceptual inventory scores (Kost *et al.*, 2009), and to understand the factors that are important in predicting physics grades early in the semester (Zabriskie *et al.*, 2019).

24.7.3 Maximum likelihood techniques

Linear regression coefficients can be determined analytically. The coefficients of other models presented in this chapter generally must be found by some search method. One of the more general and commonly applied methods is Maximum Likelihood Estimation (MLE). This method takes a probability model, for example, the model for logistic regression above, and calculates a likelihood function L from a randomly chosen initial set of model parameters and the observed data. The likelihood is the probability of the observed data given the model. To calculate the likelihood given a set of model parameters, the observed data values and the model parameters are substituted into the probability model for each observation, giving the probability that the observation occurred. The overall likelihood is the product of the probability for each observation in the data set. An algorithm iteratively modifies the model parameters until the likelihood function is maximized. The set of parameters which maximizes the likelihood function are the parameters that make the observed data most probable given the probability model. The likelihood function is used to compute the model fit function, F_{ML}. The model fit function has a chi-squared distribution, $\chi^2 = (N-1)F_{ML}$.

Many model fit statistics have been developed for MLE and have been reported in PER studies. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are information theoretic measures which characterize the information lost when applying the probability model fit to the "true" model (McElreath, 2016). Both are defined in terms of the likelihood function, as shown below. Both penalize additional parameters with BIC doing so more strongly

$$AIC = 2k - 2\, ln\, ln(L) \tag{24.14}$$

$$BIC = k\, ln\, ln(N) - 2\, ln\, ln(L) \tag{24.15}$$

where k is the number of parameters and N is the sample size. Superior models minimize AIC and BIC. Burnham and Anderson (2003) suggest a difference of 2 in AIC as significant. Raftery (1995) classifies differences in BIC: $\Delta BIC < 2$ as "weak," $2 < \Delta BIC < 6$ as "positive," $6 < \Delta BIC < 10$ as "strong," and $\Delta BIC > 10$ as "very strong." These differences are the change in AIC or BIC when the model is changed, for example, by fitting more parameters, thus changing k.

The root mean square error of approximation (RMSEA) measures badness of fit, with higher values of RMSEA representing more poorly fitting models. RMSEA 0.05 results in the rejection of the not-close-fit hypothesis. A RMSEA greater than 0.05 and less than 0.10 results in the rejection of the poor-fit hypothesis. RMSEA is defined in terms of χ^2 and the number of degrees of freedom *df, as* shown in Eq. (24.16)

$$RMSEA = \sqrt{\frac{\chi^2 - df}{(N-1)df}} \tag{24.16}$$

The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are two of a number of incremental fit statistics which compare a model to a null model. For the logistic regression above, the null model is the model containing only the intercept. Different null models are chosen for different analyses. These measures compare the χ^2 of the model fit, χ^2_{model}, to the χ^2 of the null model, χ^2_{null}. The formula for CFI is shown in Eq. (24.17). The formula for TLI is similar.

$$\mathrm{CFI} = 1 - \frac{\chi^2_{model} - df_{model}}{\chi^2_{null} - df_{null}} \tag{24.17}$$

Hu and Bentler (1999) suggest a CFI or TLI larger than 0.95 for acceptable fit and 0.97 for good fit; however, other authors suggest less restrictive criteria (Kline, 2016). Hu and Bentler (1999) also suggest using multiple fit statistics to evaluate models.

When using Multidimensional IRT to perform EFA, MLE techniques are used making MLE fit statistics available.

24.7.4 Hierarchical linear modeling

Both linear and logistic regression assume that observations are independent. This assumption is routinely violated in educational research by the intrinsic nesting of educational systems. For example, students are nested in classes which are nested in semesters and instructors, all of which are nested in institutions. All these levels of nesting produce potential correlations and violate the assumption of independence. Failure to properly account for nesting can lead to invalid conclusions. Figure 24.2 illustrates the potential for invalid conclusions which can be drawn if the nesting of data is not considered. The figure plots the FCI post-test score against the ACT percentile score for two different classes with student populations of differing levels of average academic preparation. This plot uses artificial data and is purely for illustration. In each class, the slope of the linear relation between post-test score and ACT score (the solid lines) is the same. If the data were aggregated without considering the nesting, one would incorrectly conclude that the dashed line was the correct linear relation. For this simple model where students are nested into two classrooms, one could correctly model the data by introducing a dichotomous variable into the regression, which was zero for classroom 1 and one for classroom 2. This variable would effectively adjust the intercept of the two solid regression lines. As more classrooms are added, the slopes and intercepts of each classroom should be treated as random variables.

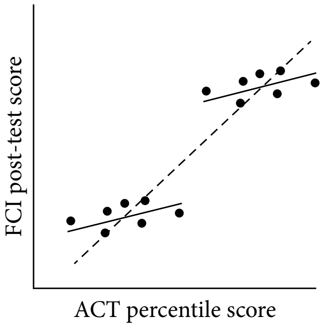


FIG. 24.2
Illustration of the effects of nested data.

For more complicated systems with many nested relations, the slopes and intercepts can be treated as random variables; the correct modeling technique for this type of data is called Hierarchical Linear Modeling (HLM). HLM divides the regression into multiple levels where lower levels are nested in higher levels. The first level, Level 1, represents the students and student-level characteristics such as pretest score. The next level, Level 2, might contain classroom-level variables such as the number of students or the experience of the instructor. Level 1 is nested within Level 2. The regression analysis would also be carried out at two levels. The Level 1 analysis is given by Eq. (24.18)

$$\text{Level 1 (student level) Post}_{ij} = \beta_{0j} + \beta_{1j} \cdot Pre_{ij} + \epsilon_{ij} \tag{24.18}$$

where i represents student i within classroom j, β_{0j} is the intercept for classroom j, β_{1j} is the slope for classroom j, and ϵ_{ij} is the residual error which is assumed to be normally distributed. The Level 2 analysis is given by Eqs. (24.19) and (24.20)

$$\text{Level 2 (classroom level intercept)}\, \beta_{0j} = \beta_2 + \beta_3 \cdot \text{classsize}_j + u_j \tag{24.19}$$

$$\text{Level 2 (classroom level slope)}\, \beta_{1j} = \beta_4 + \beta_5 \cdot \text{classsize}_j + v_j \tag{24.20}$$

where the slopes and intercepts in Level 1 are predicted by classroom-level variable $classsize_j$; u_j and v_j are random effects. For an excellent and careful reference on HLM, see West *et al.* (2015).

Van Dusen and Nissen (2019) used HLM on a large sample of CLASS data to show that nesting was important in understanding the effects of Learning Assistant programs.

24.8 STRUCTURAL EQUATION MODELING

Structural Equation Modeling (SEM) is a powerful technique widely used in general education studies that allows the combination of measurement models (latent variable models found by factor analysis) with structural models (regression models). The use of SEM is rapidly growing in PER.

Path models provide a graphical method to summarize a series of regressions. Model 1 in Fig. 24.3 shows the path model for three variables: SAT % (SAT Composite Percentile Score), a conceptual post-test score (Post-test), and physics course grade (F = 0, A = 4). The model proposes that high school academic achievement measured by SAT % affects course grades both directly and by increasing post-test scores, which then influence grades. The path model (Model 1) encodes two regressions shown in Eqs. (24.21) and (24.22). The labels A, B, and C' on the directed edges represent regression coefficients.

$$\text{Post} = \beta_0 + A \cdot \text{SAT\%} + \epsilon_1 \tag{24.21}$$

$$\text{Grade} = \beta_1 + B \cdot \text{Post} + C' \cdot \text{SAT\%} + \epsilon_2 \tag{24.22}$$

Using the path model or equivalently the regression equations, one can show that the total effect of SAT% on course grade C can be written as $C = C' + A \cdot B$ summing the two possible paths from SAT% to Grade. The total effect C can be estimated with regression Eq. (24.23).

$$\text{Grade} = \beta_2 + C \cdot \text{SAT\%} + \epsilon_3 \tag{24.23}$$

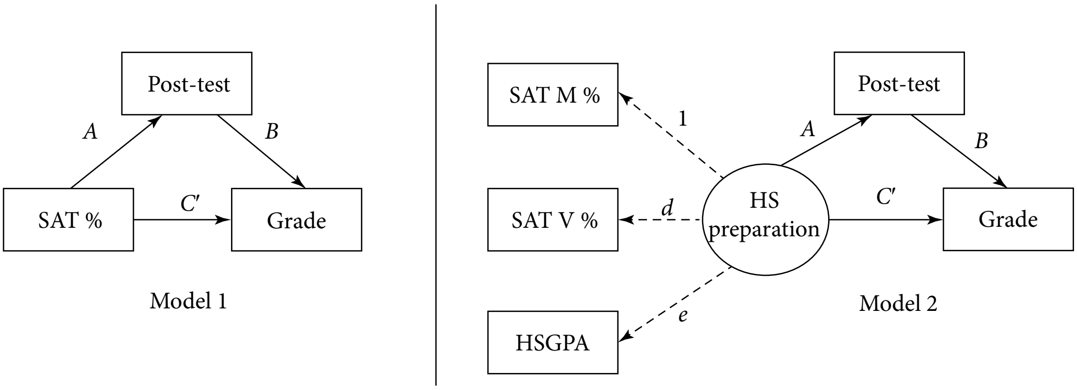


FIG. 24.3
Structural equation models.

The fraction of the effect through each path is then C'/C and $A \cdot B/C$. Model 1 represents a classically mediated relationship. In the framework of Baron and Kenny (1986), the mediation is significant if A, B, and C are significant and if $C' < C$. The framework of Baron and Kenny (1986) is insufficient for more complex mediational models and has been shown to have additional problems. In general, significant mediation is demonstrated by bootstrapping the indirect effect $A \cdot B$ and demonstrating that the 95% confidence interval does not include zero (Hayes, 2009).

Path models have been reported in many PER studies. Henderson *et al.* (2022) used path models to explore the relationship of personality facets, self-efficacy, and physics course grades. Salehi *et al.* (2019) and Stewart *et al.* (2021a) used path models to explore the relationship of demographic factors, standardized test scores, physics pretest scores, and physics course outcomes. Young and Caballero (2021) used path models to explore the role of institution size, undergraduate GPA, and physics GRE score in graduate admission in physics. Ding (2014b) used path analysis to verify the casual influences of reasoning skills and epistemological sophistication on conceptual learning outcomes.

The path models above could all be fit using traditional linear regression analysis. They can also be analyzed with SEM, which would yield the same path coefficients, but SEM uses a very different method to estimate the coefficients. Structural Equation Modeling seeks to find the set of model parameters that most closely recreates the covariance matrix of the observed variables. For Model 1 in Fig. 24.3, there are 3 observed variables; the covariance matrix is 3 by 3 matrix with $k(k+1)/2$ unique entries, where $k = 3$ is the number of variables. The covariance matrix for Model 1 has 6 unique entries, 3 variances on the diagonal and 3 covariances either above or below the diagonal (the matrix is symmetric). To fit Model 1, three regression coefficients A, B, and C' are estimated as well as the 2 variances of the regression residuals, ϵ_i, (the remaining variance accounting for the model), and the variance in *SAT*

%, which is not explained by the model; 6 total parameters. Because there are 6 unique entries in the covariance matrix and 6 parameters are fit, the model can be fit exactly. The model is said to be "just identified." SEM software uses MLE estimation and reports RMSEA, CFI, and TLI. For a just identified model, RMSEA, CFI, and TLI will be perfect (RMSEA = 0, CFI = 1, and TLI = 1) and therefore do not give meaningful information about the model fit. Note that SEM models usually do not fit the intercepts, although they can.

Structural Equation Modeling extends path analysis using only regression by allowing the inclusion of latent variables in the models. In Model 1, we used SAT composite percentile scores as a surrogate for general high school academic preparation. In SEM, a general high school preparation variable could be entered in the model as a latent variable with multiple observed indicator variables. Figure 24.3 Model 2 shows a refinement of Model 1 which introduces a latent, un-observed variable "HS Preparation" which is constructed from 3 observed variables: high school GPA, SAT mathematics percentile score, and SAT verbal percentile score. Observed variables are represented by rectangles; ovals represent latent variables. The dashed lines represent the loading of each of the observed variables on the latent variable. The modeling of the latent variable is related to a confirmatory factor analysis and can be discussed using the same terminology. The number on the line represents the factor loading. One factor loading is set to one to establish the scale; in general, SEM requires at least 3 indicator variables for each latent factor. The HS Preparation variable is then used as an independent variable in the same set of linear regressions as in Model 1.

Beyond a mathematical technique, SEM represents a modeling methodology. The paths drawn in an SEM model encode the researcher's causal hypothesis about the data. An SEM model cannot prove these hypotheses but can give evidence, if regressions or correlations are not significant, that the hypotheses are false. Weissman (2021) cautions that one should not overstate what a well- fitting SEM model implies. The causal assumptions, the directions of the arrows between nodes, are in no way supported by a finding of good model fit. The SEM model fit is invariant to reversing a line direction or to converting a line to a covariance. A directed line in an SEM model represents a causal hypothesis that one variable causes another. Two variables may simply co-vary; these relations are represented as curved bi-directional curves in SEM models (not shown in Fig. 24.3). Because SEM modeling encourages a researcher to encode causal assumptions, it also tempts one to draw causal conclusions from well- fitting models; SEM cannot establish causality but can provide evidence for later causal verification.

Beyond the concerns raised by Weissman, the SEM model fit offers substantial challenges in interpretation. The strength of SEM is the mixture of measurement and structural models, but a single fit statistic such as CFI provides a measure of the combined fit of both the measurement and the structural model. This opens the possibility that a model with excellent CFI represents a very good measurement model and a poor structural model. If latent variables are included in a model, the measured variables which load on these variables often carry much of the variance of the model. In this case, the measurement model alone may dominate the calculation of incremental fit statistics such as CFI and TLI. To avoid inaccurate conclusions about the structural model fit, one should first estimate the measurement model alone and then estimate the improvement the structural model makes on the measurement model (Schumacker and Lomax, 2016).

Structural equation modeling has been used in a number of PER studies. Li and Singh (2022) applied SEM to explore changes in students' self-beliefs over a two-course physics sequence. Salehi *et al.* (2019) used SEM to fit path models examining the relation of demographics, standardized test scores, pretest scores, and final exam grades. Li and Singh (2021) used SEM to investigate the relation of gender, self-efficacy, and identity in a physics course. Lock *et al.* (2019) used SEM to explore factors involved in physics identity development and career choice.

While not currently used in PER, there is an exciting strand of research advanced by Pearl (2009) that is popular in other fields such as epidemiology, which claims to use SEM models to make causal inferences.

For the classic discussion of SEM see Kline (2016) and for a more detailed discussion see Schumacker and Lomax (2016).

24.9 CLUSTER ANALYSIS

Cluster Analysis (CLA), like PCA, is a dimensional reduction technique which seeks to explain a set of data with a more parsimonious set of variables; however, the two methods accomplish this in a completely different manner. Factor analysis estimates a set of latent traits, factors, where each student has some level of each factor. Cluster analysis seeks to divide the set of students into groups, clusters, where members of one cluster are similar to each other but different from members of another cluster. Cluster analysis is intrinsically an exploratory technique; there is no provision to provide a theoretical cluster model. All cluster algorithms require some distance metric; many such metrics exist. For conceptual inventory or other numerical data, a Euclidean metric is common. The Euclidean metric treats each data record as a vector in a k-dimensional space. For conceptual inventories, this would be a vector of zeros and ones representing whether each item was answered correctly. The distance between two response vectors is calculated by extending the Pythagorean theorem to the k-dimensional space. For a discussion of cluster analysis, see Xu and Wunsch II (2009).

Multiple algorithms for CLA have been created. Two common methods that have been applied in PER are agglomerative hierarchical clustering (AHC) and k-means clustering. In AHC, all data records (response vectors) begin as their own cluster. This initial set of clusters is iteratively reduced by joining the two nearest clusters using the distance metric to form one larger cluster. If more than two clusters are equidistant at the nearest distance, two of these clusters are joined at random. This continues until only one cluster containing all data records remains. A dendrogram, as shown in Fig. 24.5, summarizes the process. The leaves represent individual student response vectors. The vertical axis plots the dissimilarity between the two clusters merged; in this case, the dissimilarity is the distance between the cluster centers. To identify the optimal number of clusters, a scree plot like that shown in Fig. 24.4

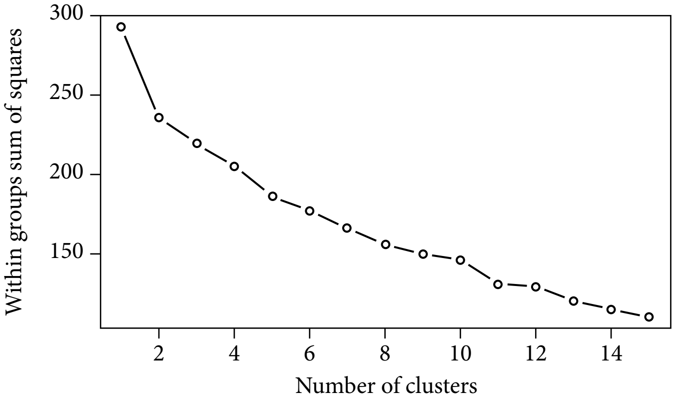


FIG. 24.4
Scree plot for cluster analysis for 50 student responses to the FCI.

can be used. The scree diagram plots the remaining error sum of squares against the number of clusters. A weak "knee" is visible at two clusters. One can also examine the dendrogram for the set of mergers with the largest change in dissimilarity. For the data plotted in Fig. 24.5, the dendrogram also suggests that two clusters are optimal. The red rectangles on the figure separate the dendrogram into these two sets of leaves. AHC does not allow the algorithm to overcome mistakes in clustering (it can find local minima). The k-means algorithm is a stochastic algorithm that can partially overcome this problem. In k-means, one begins by selecting the target number of clusters. A set of random cluster centers is then generated. Each data point is added

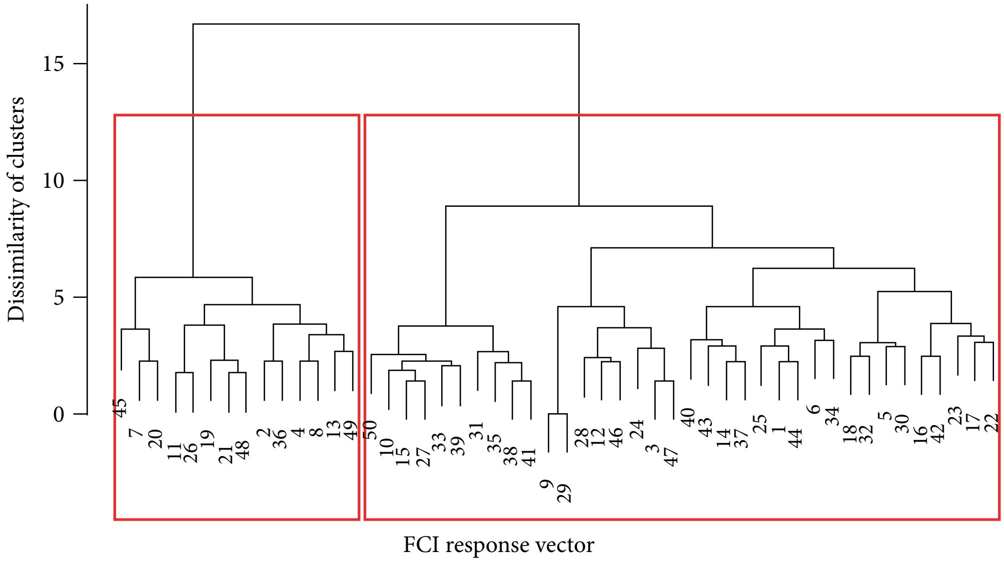


FIG. 24.5
Cluster dendrogram for 50 student responses to the FCI.

to its nearest cluster, the average center of the data points is calculated and becomes the new center of the cluster, the data points are then moved to the new nearest cluster, and the process continues until a convergence criterion is met.

Ding and Beichner (2009) discussed CLA as one of the five techniques used for the quantitative analysis of multiple-choice instruments in PER. Battaglia *et al.* (2019) provided a general exploration of clustering techniques in PER. Cluster analysis has been used in a number of PER studies. Klein *et al.* (2021) applied cluster analysis to understand patterns in eye tracking measurements of kinematic graphs. Springuel *et al.* (2007) applied CLA to understand student reasoning about 2D kinematic graphs. Stewart *et al.* (2012) used CLA to examine the progression of the consistency of students answering isomorphic questions. Fazio and Battaglia (2019) applied CFA to understand response patterns to the FCI. Stewart *et al.* (2021b) applied CFA to show that across 3 institutions with substantially different FCI post-test scores, a 3-cluster model with similar centroids was optimal for explaining latent traits involving Newtonian and non-Newtonian thinking identified by IRT.

24.10 NETWORK ANALYSIS

A network is formed by a collection of nodes connected by edges to form a graph. The edges may be directed or un-directed. The edges may also be weighted to indicate some features of the interaction. Note that the term edge comes from graph theory on which network analysis is based. Often the nodes represent people and the edges some interaction between people, such as membership in a lab group; however, the nodes could also be responses to items in an RBI and the edge correlations of those responses. Brewe (2018) provides a summary of network analysis in PER.

24.10.1 Social Network analysis

The most common application of network analysis in PER is to characterize social structures arising in the physics classroom or between physics students. In these networks, the nodes are students or instructors, and the edges represent some form of social interaction. These networks and the metrics available in social network analysis (SNA) have been used to characterize learning environments (Traxler *et al.*, 2020; Commeford *et al.*, 2021; and Sundstrom *et al.*, 2022), to predict class achievement (Bruun and Brewe, 2013; and Vargas *et al.*, 2018) or retention to a degree program (Forsman *et al.*, 2014; and Zwolak *et al.*, 2017). SNA has also been used to explore anxiety and self-efficacy in physics (Dou *et al.*, 2016; and Dou and Zwolak, 2019), gender disaggregated interactions in physics lab groups (Sundstrom *et al.*, 2022), and to study the conceptual change (Bodin, 2012; and Bruun *et al.*, 2019).

24.10.2 Module analysis

Module analysis applies network analysis to a network formed of the responses to multiple-choice instruments. These analyses identify groups of responses preferentially selected together by students. For incorrect responses, these may represent consistently applied misconceptions. For correct responses, these groups may represent isomorphic items. Each response to the instrument becomes a node; for example, FCI item 8 response A would become node 8A. The nodes are connected with weighted edges. Different forms of module analysis compute the edge weights as the following: the number of times the item is selected together (MAMCR), the correlation between the responses (MMA), and the partial correlation between the responses correcting for instrument score (MMA-P). Brewe *et al.* (2016) developed module analysis as module analysis for multiple choice responses (MAMCR) using the number of times two responses were selected together as the edge weight. Wells *et al.* (2019) attempted to apply MAMCR to a large sample and found that the algorithm was not productive for larger samples. Wells *et al.* developed Modified Module Analysis (MMA) using the correlation between responses as the edge weight to allow the analysis of large data sets. Neither MAMCR nor MMA could analyze networks including correct responses; Yang *et al.* (2020) extended MMA to include correct responses by introducing Modified Module Analysis Partial (MMA-P); MMA-P weighted the edges with the partial correlation correcting for total instrument score. Module analysis has been applied to the FCI (Brewe *et al.*, 2016; Wells *et al.*, 2020; and Yang *et al.*, 2020), the FMCE (Yang *et al.*, 2019; and Wells *et al.*, 2020), the CSEM (Wheatley *et al.*, 2021) as well as a conceptual assessment of quantum mechanics (Wells *et al.*, 2021).

24.11 MACHINE LEARNING

The explosion of computing power and the amount of data collected by both corporate and educational entities has resulted in the development of new technologies generally classified as machine learning. These involve both new algorithms and new philosophies for applying those algorithms that could be potentially transformative for PER and education in general. Many general reviews of these methods are available (Müller and Guido, 2016). Some algorithms such as logistic regression applied in machine learning are in common use in PER; others such as neural networks are not. Machine learning brings new conceptual tools to quantitative analysis, including the methodology of splitting data sets into test and training data sets, so that models can be evaluated on data not used to build the models. Machine learning also often relies on the concept of classification, where the algorithms are used to predict some outcome; machine learning provides a wealth of techniques to characterize the classification process. Some of these ideas may inform general quantitative research in PER (Aiken *et al.*, 2021).

Machine learning applied to educational systems is part of a broad initiative in general education research called educational data mining that has been ongoing for 20 years. Many general reviews of the status of this field are available (Romero and Ventura, 2010; and Pena-Ayala, 2014). These studies often predict either student classroom outcomes (grades or final exam scores) or general student retention. Machine learning or more generally artificial intelligence has many additional applications, including intelligent tutoring systems and automated grading of assignments (Chen *et al.*, 2020).

Within PER, machine learning has been used to predict physics student retention (Aiken *et al.*, 2019) and student success in introductory physics (Zabriskie *et al.*, 2019). Beyond the prediction, natural language processing using machine learning algorithms has also been used to classify open response answers to physics instruments (Wilson *et al.*, 2022).

24.12 CONCLUSION

A rich spectrum of quantitative research methods have been applied in PER to understand student learning in physics classes, the properties of instruments which measure that learning, the experiences and outcomes of students in those classes, and the interaction of different social groups within those classes. This chapter has surveyed only a few of the many methods which have been productive in PER. New methods are continuously being brought to the field and continue to expand our understanding of physics education.

ACKNOWLEDGMENTS

We thank Christopher Wheatley for his valuable input into the network analysis section of the work.

REFERENCES

Adams, W. *et al.*, Phys. Rev. Phys. Educ. Res. **2**, 010101 (2006).
Aiken, J. *et al.*, Phys. Rev. Phys. Educ. Res. **17**, 020104 (2021).
Aiken, J. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(1), 010128 (2019).
Andrade, C., Indian J. Psychol. Med. **41**(1), 99–100 (2019).
Bao, L., Am. J. Phys. **74**(10), 917 (2006).
Baron, R. and Kenny, D., J. Personal. Soc. Psychol. **51**(6), 1173 (1986).
Barthelemy, R. S. and Knaub, A. V., Phys. Rev. Phys. Educ. Res. **16**, 010133 (2020).
Barthelemy, R. *et al.*, Phys. Rev. Phys. Educ. Res. **18**, 010124 (2022).
Battaglia, O. *et al.*, Phys. Rev. Phys. Educ. Res. **15**, 020112 (2019).
Bock, R., Psychometrika **37**(1), 29–51 (1972).
Bodin, M., Phys. Rev. Phys. Educ. Res. **8**, 010115 (2012).
Brewe, E., *Getting Started in PER*, 4th ed. (PER Central, College Park, MD, 2018), Vol. 2, see https://www.per-central.org/items/detail.cfm?ID=14725.
Brewe, E. *et al.*, Phys. Rev. Phys. Educ. Res. **12**, 020131 (2016).
Bruun, J. and Brewe, E., Phys. Rev. Phys. Educ. Res. **9**, 020109 (2013).
Bruun, J. *et al.*, Int. J. Res. Method Educ. **42**(3), 317–339 (2019).
Burnham, K. and Anderson, D., *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer-Verlag, New York, NY, 2003).
Calin-Jageman, R. J. and Cumming, G., Am. Stat. **73**(Supp1.), 271–280 (2019).
Canty, A. and Ripley, B., *Boot: Bootstrap R (S-Plus) Functions.* R package version 1.3-20 (2017).
Chabay, R. and Sherwood, B., AAPT Announcer **27**, 96 (1997).
Chen, L. *et al.*, IEEE Access **8**, 75264–75278 (2020).
Christman, E. *et al.*, *AIP Conference Proceedings, 2020 PERC Proceedings [Virtual Conference, July 22–23]* (AIP, College Park, MD, 2020), see https://www.compadre.org/per/items/detail.cfm?ID=15463.
Cohen, J., *Statistical Power Analysis for the Behavioral Sciences* (Academic Press, New York, NY, 1977).
Cohen, J. *et al.*, *Applied Multiple Regression/Correlation Analysis for the Behavioral Science*, 3rd ed. (Routledge, New York, NY, 2003).

Coletta, V. *et al.*, Phys. Rev. Phys. Educ. Res. **3**(1), 010106 (2007).
Coletta, V. and Steinert, J., Phys. Rev. Phys. Educ. Res. **16**(1), 010108 (2020).
Commeford, K. *et al.*, Phys. Rev. Phys. Educ. Res. **17**, 020136 (2021).
Crocker, L. and Algina, J., *Introduction to Classical and Modern Test Theory* (Holt, Rinehart and Winston, New York, NY, 1986).
Cumming, G., Psychol. Sci. **25**(1), 7 (2014).
Davison, A. and Hinkley, D., *Bootstrap Methods and Their Applications* (Cambridge University Press, Cambridge, 1997).
Ding, L., Phys. Rev. Phys. Educ. Res. **10**(1), 010105 (2014a).
Ding, L., Phys. Rev. Phys. Educ. Res. **10**(2), 023101 (2014b).
Ding, L., Phys. Rev. Phys. Educ. Res. **15**(2), 020101 (2019).
Ding, L., *Advances in Applications of Rasch Measurement in Science Education* (Springer Nature, New York, NY, 2022).
Ding, L. and Beichner, R., Phys. Rev. Phys. Educ. Res. **5**, 020103 (2009).
Ding, L. *et al.*, Phys. Rev. Phys. Educ. Res. **2**(1), 010105 (2006).
Ding, L. and Lui, X., *Getting Started in PER* (PER, 2012), Vol. 2, see https://www.per-central.org/items/detail.cfm?ID=12601.
Docktor, J. and Mestre, J., Phys. Rev. Phys. Educ. Res. **10**(2), 020119 (2014).
Dou, R. *et al.*, Phys. Rev. Phys. Educ. Res. **12**, 020124 (2016).
Dou, R. and Zwolak, J., Phys. Rev. Phys. Educ. Res. **15**, 020105 (2019).
Douglas, K. A. *et al.*, Phys. Rev. Phys. Educ. Res. **10**, 020128 (2014).
Eaton, P. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020133 (2019a).
Eaton, P. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(1), 010102 (2019b).
Eaton, P. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020151 (2019c).
Eaton, P. and Willoughby, S., Phys. Rev. Phys. Educ. Res. **16**(1), 010106 (2020).
Efron, B., *Break- Throughs in Statistics* (Springer, New York, NY, 1992), pp. 569–593.
Ellis, P., *The Essential Guide to Effect Sizes* (Cambridge University Press, Cambridge, 2010).
Fazio, C. and Battaglia, O., Int. J. Sci. Math. Educ. **17**(8), 1497–1517 (2019).
Forsman, J. *et al.*, Stud. Higher Educ. **39**(1), 68 (2014).
Freeman, S. *et al.*, Proc. Natl. Acad. Sci. U.S.A. **111**(23), 8410 (2014).
Gelman, A. *et al.*, *Regression and Other Stories* (Cambridge University Press, Cambridge, 2021).
Greenland, S. *et al.*, Eur. J. Epidemiol. **31**(4), 337–350 (2016).
Hake, R., Am. J. Phys. **66**, 64 (1998).
Halloun, I. and Hestenes, D., Am. J. Phys. **53**(11), 1056 (1985a).
Halloun, I. and Hestenes, D., Am. J. Phys. **53**(11), 1043 (1985b).
Han, J. *et al.*, Phys. Rev. Phys. Educ. Res. **11**, 010112 (2015).
Hansen, J. and Stewart, J., Phys. Rev. Phys. Educ. Res. **17**, 020139 (2021).
Hayes, A., Commun. Monogr. **76**(4), 408–420 (2009).
Hazari, Z. *et al.*, Phys. Rev. Phys. Educ. Res. **9**, 020115 (2013).
Heller, P. and Huffman, D., Phys. Teach. **33**(8), 503 (1995).
Henderson, C. *et al.*, Phys. Rev. Phys. Educ. Res. **8**(2), 020104 (2012).
Henderson, R. *et al.*, Phys. Rev. Phys. Educ. Res. **18**, 010143 (2022).
Henderson, R. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(2), 020103 (2018).
Hestenes, D. and Halloun, I., Phys. Teach. **33**(8), 502 (1995).
Hestenes, D. and Jackson, J., Table II for the Force Concept Inventory (revised from 081695r) (2010), see http://modeling.asu.edu/R&E/FCI-RevisedTable-II_2010.pdf (last accessed March 17, 2019).
Hestenes, D. *et al.*, Phys. Teach. **30**, 141 (1992).
Hu, L. and Bentler, P., Struct. Eq. Model. **6**(1), 1–55 (1999).
Huffman, D. and Heller, P., Phys. Teach. **33**, 138 (1995), see https://www.compadre.org/per/items/Load.cfm?ID=2797.
Klein, P. *et al.*, Phys. Rev. Phys. Educ. Res. **17**, 013102 (2021).
Kline, R., *Principles and Practices of Structural Equation Modeling, Fourth Edition* (Guilford Publications, New York, NY, 2016).
Kost, L. *et al.*, Phys. Rev. Phys. Educ. Res. **5**, 010101 (2009).
Kubsch, M. *et al.*, Pract. Assess. Res. Eval. **26**(1), 4 (2021).
Lasry, N. *et al.*, Am. J. Phys. **79**(9), 909–912 (2011).
Li, Y. and Singh, C., Phys. Rev. Phys. Educ. Res. **17**, 010143 (2021).
Li, Y. and Singh, C., Phys. Rev. Phys. Educ. Res. **18**, 010142 (2022).
Lock, R. *et al.*, Phys. Rev. Phys. Educ. Res. **15**, 020137 (2019).
Madsen, A. *et al.*, Am. J. Phys. **85**(4), 245 (2017).
Maloney, D. *et al.*, Phys. Educ. Res., Am. J. Phys. Suppl. **69**(S1), S12 (2001).
Marx, J. and Cummings, K., Am. J. Phys. **75**(1), 87 (2007).
McDermott, L. and Redish, E., Am. J. Phys. **67**(9), 755 (1999).
McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (CRC Press, Taylor & Francis Group, Boca Raton, FL, 2016).
Meltzer, D. and Thornton, R., Am. J. Phys. **80**(6), 478 (2012).
Müller, A. and Guido, S., *Introduction to Machine Learning with Python: A Guide for Data Scientists* (O'Reilly Media, Boston, MA, 2016).

Nissen, J. *et al.*, Phys. Rev. Phys. Educ. Res. **15**, 020106 (2019).
Nissen, J. *et al.*, Phys. Rev. Phys. Educ. Res. **14**(1), 010115 (2018).
Osborn Popp, S. *et al.*, *2011 American Educational Research Association Conference* (American Education Research Association, Washington, DC, 2011), see https://www.aera.org.
Pearl, J., *Causality* (Cambridge University Press, Cambridge, 2009).
Pena-Ayala, A., Expert Syst. Appl. **41**(4), 1432–1462 (2014), see https://www.sciencedirect.com/science/article/abs/pii/S0957417413006635.
PhysPort, Physport (2022), see https://www.physport.org. Accessed 8/14/2022.
Planinic, M. *et al.*, Phys. Rev. Phys. Educ. Res. **6**, 010103 (2010).
Pollock, S., *2008 Physics Education Research Conference Proceedings, Volume 1064*, edited by C. Henderson *et al.* (AIP Publishing, New York, 2008), pp. 171–174, see https://www.compadre.org/per/items/detail.cfm?ID=8109.
Raftery, A., Sociol. Methodol. **25**, 111–163 (1995).
Ramlo, S., Am. J. Phys. **76**(9), 882–886 (2008).
Redish, E. *et al.*, Am. J. Phys. **66**(3), 212–224 (1998).
Romero, C. and Ventura, S., IEEE Trans. Syst. Man Cybernet. Part C **40**(6), 601–618 (2010), see https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5524021.
Salehi, S. *et al.*, Phys. Rev. Phys. Educ. Res. **15**, 020114 (2019).
Sass, D. A. and Schmitt, T. A., Multivar. Behav. Res. **45**(1), 73–103 (2010), see https://www.researchgate.net/publication/232890796_A_Comparative_Investigation_of_Rotation_Criteria_Within_Exploratory_Factor_Analysis.
Schroeder, C. *et al.*, J. Res. Sci. Teach. **44**(10), 1436 (2007).
Schumacker, R. and Lomax, R., *A Beginners Guide the Structural Equation Modeling*, 4th ed. (Routledge, New York, NY, 2016).
Scott, T. and Schumayer, D., Phys. Rev. Phys. Educ. Res. **11**, 020134 (2015).
Scott, T. *et al.*, Phys. Rev. Phys. Educ. Res. **8**(2), 020105 (2012).
Semak, M. *et al.*, Phys. Rev. Phys. Educ. Res. **13**, 010103 (2017).
Smith, T. and Bendjilali, N., Phys. Rev. Phys. Educ. Res. **18**, 010133 (2022).
Smith, T. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010107 (2020).
Springuel, R. *et al.*, Phys. Rev. Phys. Educ. Res. **3**(2), 020107 (2007).
Stewart, J. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(1), 010107 (2021a).
Stewart, J. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(1), 010122 (2021b).
Stewart, J. *et al.*, Phys. Rev. Phys. Educ. Res. **8**(2), 020112 (2012).
Stewart, J. and Stewart, G., Phys. Teach. **48**(3), 194 (2010).
Stewart, J. *et al.*, Phys. Rev. Phys. Educ. Res. **14**, 010137 (2018).
Sundstrom, M. *et al.*, Phys. Rev. Phys. Educ. Res. **18**, 010102 (2022).
Thornton, R. *et al.*, Phys. Rev. Phys. Educ. Res. **5**(1), 010105 (2009).
Thornton, R. and Sokoloff, D., Am. J. Phys. **66**(4), 338 (1998).
Traxler, A. *et al.*, Phys. Rev. Phys. Educ. Res. **14**, 010103 (2018).
Traxler, A. *et al.*, Phys. Rev. Phys. Educ. Res. **16**, 020129 (2020).
van der Linden, W., *HandBook of Item Response Theory, Volume 1* (CRC Press, Taylor & Francis Group, New York, 2016), pp. 13–30.
Van Dusen, B. and Nissen, J., Phys. Rev. Phys. Educ. Res. **15**, 020108 (2019).
Vargas, D. *et al.*, Phys. Rev. Phys. Educ. Res. **14**, 020112 (2018).
Von Korff, J. *et al.*, Am. J. Phys. **84**(12), 969 (2016).
Wang, J. and Bao, L., Am. J. Phys. **78**(10), 1064–1070 (2010).
Weissman, M. B., Phys. Rev. Phys. Educ. Res. **17**, 020118 (2021).
Wells, J. *et al.*, Phys. Rev. Phys. Educ. Res. **15**, 020122 (2019).
Wells, J. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(1), 010121 (2020).
Wells, J. *et al.*, Phys. Rev. Phys. Educ. Res. **17**, 020113 (2021).
West, B. *et al.*, *Linear Mixed Model: A Practical Guide Using Statistical Software* (CRC Press, Boca Raton, FL, 2015).
Wheatley, C. *et al.*, Phys. Rev. Phys. Educ. Res. **17**, 010102 (2021).
Wilson, J. *et al.*, Phys. Rev. Phys. Educ. Res. **18**, 010141 (2022).
Xiao, Y. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020149 (2019).
Xu, R. and Wunsch, II, D., *Clustering* (Wiley, Hoboken, NJ, 2009).
Yang, J. *et al.*, Phys. Rev. Phys. Educ. Res. **16**, 010124 (2020).
Yang, J. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020141 (2019).
Yasuda, J. *et al.*, Phys. Rev. Phys. Educ. Res. **18**, 010112 (2022).
Young, N. and Caballero, M., Phys. Rev. Phys. Educ. Res. **17**, 010144 (2021).
Yu, C. and Yao, W., Commun. Statistics-Simul. Comput. **46**(8), 6261–6282 (2017).
Zabriskie, C. and Stewart, J., Phys. Rev. Phys. Educ. Res. **15**, 020107 (2019).
Zabriskie, C. *et al.*, Phys. Rev. Phys. Educ. Res. **15**, 020120 (2019).
Zwickl, B. M. *et al.*, Phys. Rev. Phys. Educ. Res. **10**, 010120 (2014).
Zwolak, J. *et al.*, Phys. Rev. Phys. Educ. Res. **13**, 010113 (2017).

CHAPTER 25 QUALITATIVE METHODS IN PHYSICS EDUCATION RESEARCH

Valerie K. Otero, Danielle Boyd Harlow, and David E. Meltzer

Otero, V. K., Harlow, D. B., and Meltzer, D. E., "Qualitative methods in physics education research," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 25-1–25-32.

25.1 INTRODUCTION

PER researchers often decide to engage in qualitative studies in order to gather wider varieties of data and/or data of greater depth than may be accessible in purely quantitative studies. Qualitative studies use different types of data and analysis methods than quantitative studies, but like quantitative studies, they use evidence to make and support claims about physics learning and teaching.

In this chapter, we provide an introduction to strategies and procedures for collecting and analyzing qualitative data and discuss other aspects of qualitative research such as theoretical framing. There are many different traditions of qualitative research, for example: Glaser and Strauss, 1967; Psathas, 1973; Lemke, 1990; Strauss and Corbin, 1990; LeCompte *et al.*, 1993; Riessman, 1993; and Gee, 2004 (also see Denzin and Lincoln, 2000; and Green *et al.*, 2006). Each of these traditions of qualitative research has some things in common with the others, but each is replete with its own language, jargon, and terminology. In recognition of these challenges, we provide in this chapter a generic approach to qualitative research that will provide researchers with an overview of the field. We illustrate general methods with specific examples drawn from the PER literature, some of which are discussed in considerable depth, to provide a clear sense of how general methods are applied in real-world research settings. *It is important to emphasize that we do not attempt to provide a comprehensive review of qualitative research methods used in PER*; that would be impractical, given the truly enormous variety of methods that are used—described in many hundreds of research papers—and the length constraints imposed on the chapters in this volume. Nor do we claim that the examples we provide are by any measure the "best" illustrations of the cited methods. The examples are drawn both from well-known

https://doi.org/10.1063/9780735425514_025

older studies and—somewhat at random—from the large and rapidly growing recent literature. A comprehensive search through the literature was simply not a practical option that we could consider, given that the variety of qualitative methods is nearly limitless.

This chapter is largely an abridged and revised version of Otero and Harlow (2009), with many additional examples added from the more recent PER literature.

25.2 OVERVIEW

Inductive analysis is the type of analysis in which the researcher seeks to derive trends, concepts, themes or a model through multiple reads of the data. In a *deductive* approach, the researcher typically applies *a priori* criteria or assumptions—determined in advance—and seeks to determine whether the data are consistent with them. Most qualitative analyses involve a little bit of both. In this article, we will spend most of the time discussing the inductive analytic method. There are at least three reasons for performing inductive analyses (Thomas, 2006):

1. To condense extensive and varied raw text data into a brief summary format.
2. To establish clear links between the research objectives and the summary findings derived from the raw data, and to ensure that these links are both transparent (demonstrable to others) and defensible (justifiable given the objectives of the research) (Gee, 2004); and
3. To develop a model or theory about the underlying structure of experiences or processes that are evident in the text data (i.e., in the raw text).

The terms “model” and “theory” in qualitative research are typically used very differently from the way they are used in physics and quantitative research. In qualitative research, the terms “theory” and “model” often refer to a *generalized description* of the data, that is, a shorthand way of summarizing and conceptualizing the salient features of the data. Before we begin to discuss how qualitative data are collected, analyzed and interpreted, it is useful to take a look at some of the differences between quantitative and qualitative data analysis (Cook and Reichardt, 1979).

Qualitative research tends to be inductive in that the researcher first explores the data generated by the study, and from this works toward a general model that can describe or attempt to explain the data. In quantitative research, by contrast, one typically applies *a priori* criteria or assumptions and seeks to determine whether the data are consistent with them. Qualitative research tends to be subjective rather than objective in the following senses. First, qualitative researchers attempt to describe the world from the perspective of the people studied and do not usually attempt to generalize findings to all members of a specific group. Second, researchers bring with them specific ideas regarding learning physics, and about how people do (or should) approach learning physics. Good qualitative researchers acknowledge the many subjectivities or tacit assumptions that tend to guide their actions and interpretations. These biases and tacit theory impact all aspects of the research process. The qualitative researcher strives to report a contextually-bound situation in a way that is useful for other researchers but may not be generalizable beyond the context of the study. (One does presume that similar contexts will arise elsewhere, or that at least some aspects of the study will be relevant to physics learning in other contexts.)

Qualitative researchers tend to take an anthropological world view in the sense that they usually first investigate "what is out there" and then begin to create models that can describe the actors, their actions, and their interactions. Although research in natural science is also often like this, it is safe to say that in qualitative research, one typically does not attempt to develop theories that can lead to accurate predictions. In many cases, the qualitative researcher does not attempt to control variables, preferring instead to observe the context as it is. However, there are times when a researcher wants to use qualitative techniques to conduct studies that are quasi-experimental. If so, the researcher will need to take some measures to control variables, although this looks very different than in quantitative research. Typically, in qualitative research, the researcher establishes controls by defining a "matched sample" of participants. A matched sample is a group of individuals who have similar backgrounds as the study group but differ in ways that are relevant to the study. For example, a researcher might wish to study the effects of a retention program on female graduate students. The researcher can establish a control group by finding female graduate students of similar age, GPA, and science backgrounds who did not take part in the retention program.

Descriptive or explanatory models created through qualitative research only claim to describe or explain a particular situation with particular subjects at a particular time. Replication may be difficult due to the contextual nature of these studies, that is, to the myriad of uncontrolled variables; it is not usually possible to acquire the very large sample sizes that might lead to easier generalizability. Qualitative research can provide what is known as a "thick description" of a situation and the actors that shape it, implying the inclusion of a wide variety of potentially significant details in the research report that may turn out to be valuable for analysis. Specifically, this term implies attention to the multiple and intersecting layers of inference, implication, and interpretation that are needed to draw meaning from the observation of complex social interactions (Geertz, 1973). Overall, qualitative research is different from but complementary to quantitative research and both have their costs and benefits. In the sections that follow, we hope to illustrate the many benefits of qualitative research and demonstrate how it has contributed to the PER literature base.

25.3 RESEARCH QUESTIONS AND STUDY DESIGN

A researcher often begins a qualitative study with a research idea rather than with specific research questions. The research idea represents a broad sense of what the researcher is interested in knowing more about. The *research questions* are more specific than the research idea and have to do with what can actually be observed or measured in a particular research context; the research questions are sometimes changed throughout the process of data collection and analysis (Peshkin, 1985; and Strauss and Corbin, 1990).

See the list below for questions that researchers might ask themselves when beginning to develop research questions.

- **What is the research idea?** (This does not refer to specific research questions; this is instead just a "big idea" that could be of interest to the broader community (or some fraction of it). For example, "How do people develop physics knowledge?")
  This question can probably not be answered directly, but light can be shed on the question through a carefully designed research study.
- **Do I have an implicit hypothesis? What might it be?**
- **Have I made specific assumptions? What are they?**
- **What are the potential population(s) and context(s) for investigation:**
- **Why might this population/context be a good population/context for the study?**
- **What are some of the things I might be able to "measure" (or describe) that could be relevant to the research idea?** ("Measure" could be anything like changes in beliefs, actions, scores on an instrument, time spent in sense-making activity, conceptions…).
- **What are some problems I think I might run into** (logistically, empirically, theoretically, applicability, generalizability, confounding factors)**?**
- **What are some possible ways of dealing with these problems? Why might these be helpful?**
  Theoretically (if applicable) (For example, what theory or framework can help me work through distinctions between groups and individuals, between concepts and beliefs, between observables and non-observables, etc…?)
  Empirically (if applicable) (Do I need to shift the focus of the question so that what I'm interested in knowing can fall out of the analysis rather than being the question I ask directly? Do I need a control group or do I need to collect baseline data?)
- **What things might I still need to learn to carry out a study like this?** (research methods, research literature, theory, more about the population/context of interest…)

In the first question, the "research idea" represents the broad issue to which the researcher hopes the findings can contribute; for example, researchers might want to know *why* courses that use interactive engagement yield higher learning gains than those that use traditional lecture methods, or to investigate the costs and benefits of using computer simulators vs laboratory apparatus in the physics classroom. Both of these are "big" questions and would be difficult to answer in a reasonable amount of time and with the resources and settings available to a single researcher. However, as we see below, the researcher can design some questions and a study that can shed light on these bigger issues.

As the researcher works to reveal tacit assumptions, the problem framing begins to become clear and questions begin to come into focus. For example, a researcher who makes the assumption that success in solving conceptual physics problems is largely related to the extent to which the student understands the physics might begin by comparing the academic preparation in science content knowledge of students who perform very well on these types of problems to those who do not. However, another researcher might assume that social and cultural issues are very strongly associated with success in classroom problem-solving. This researcher might then investigate what students actually think they are being asked to do when they are confronted with a physics problem, as well as how they react to those expectations, and only later check to see how this sample of students actually performed on conceptual exams. Both studies will yield results that will contribute to the community's understanding of differences in students' physics performance, but each study has different emphases and will require different methods.

Although it is not possible to control all relevant variables, researchers should focus their research questions on the specific population in which they are most interested, for example, non-physics majors or physics majors? Physics students or physics teachers? A whole different set of questions can be asked about undergraduate students in a course for non-majors than about physics graduate students in their final year of graduate school. In addition, researchers have to be realistic about the population(s) to which they have access.

A research question asks what things might be measurable or observable, given the research idea, the research contexts, and the potential study populations available to the researcher. The research question must be focused and suggest some type of measurement. The term "measurement" here should be defined broadly. For example, the researcher might have access to the recitation sections in which the University of Washington *Tutorials for Introductory Physics* are implemented. The researcher could "measure" *qualitatively* how the students interact with the tutorial worksheets, and the ways in which they interact with one another, by video recording them as they work through a tutorial. One can imagine that the researcher could observe certain trend-like behaviors within the group or among individuals who make up the group (after watching the video recording several times). These trends represent signals or measurements that can later be compared to one another or to similar measurements that used different groups in the same context or different contexts.

Once the researchers have decided on a research idea and developed initial research questions (even if these do not end up being the final research questions), they must determine how, when, and what type of data to collect to help answer these questions.

25.3.1 Collecting data

Qualitative studies involve the collection of descriptive non-numerical data. This data collection is often limited by the time and expense necessary to collect and analyze it. Ideally, enough data has been collected when additional data does not reveal anything new. Realistically, external constraints such as the end of the semester or end of funding will impact decisions about when data collection ends. The number of participants in qualitative studies is necessarily much smaller than that in quantitative studies. In some cases, researchers may choose to study only one individual or context. Other studies may look across a handful or dozens of individuals or contexts. The fact that qualitative research often deals with small numbers of participants means that it is often desirable to purposefully select participants who will provide the greatest insight into the research problem. This differs from the random selection of participants often used in quantitative studies that are used for generalization across populations. When purposefully sampling, the researcher carefully chooses subjects according to some pre-selected criteria, for example: extreme cases or, instead, typical cases; maximum-variation samples or, instead, samples that require participants to meet specified narrow criteria (Patton, 1990). For example, if the study intends to understand the best practices in solving physics problems, one may choose to recruit only students who are in the top quarter of the class or who score the highest on a particular exam. In contrast, another study of problem-solving skills may choose to sample by selecting participants who represent the widest range of experiences or abilities with the selected topic.

In qualitative research, many types of data are collected. For example, classroom observation data may help a researcher understand what an instructor actually does while teaching a physics course, whereas an interview with the instructor may shed light on the instructors' goals and intentions. Here, we describe the primary methods by which qualitative data are collected and highlight some of the advantages and concerns specific to each method.

25.4 INTERVIEWS

Interviews are an important method for collecting data. Critical steps taken prior to and during an interview include (1) determining the type of interview, (2) developing an interview protocol, and (3) conducting and recording the interview. The details of the interview design will depend critically on the *purposes* and research aims of the interview; these will be discussed and illustrated with examples from the research literature.

25.4.1 Types of interviews

- Individual vs Focus Group
- Unstructured vs Semi-structured vs Structured
- Think-aloud vs Stimulated-recall vs Artifact-based interview

Individual vs *Focus Group*. In an individual interview, one interviewer asks questions of one interviewee. Individual interviews allow researchers to attribute ideas and thoughts to a single participant. Focus group interviews involve more than one interviewee. Focus groups have been used extensively in marketing research because participants are sometimes triggered by other participants' responses, thus generating a wider range of responses (Stewart and Shamdasani, 1990).

Unstructured vs *Semi-structured* vs *Structured Interviews*. Interviews vary in the degree to which the interview as a whole and the individual questions are pre-determined, narrow, and prescriptive, that is, constraining in advance the scope of the interviewee's responses; the degree of constraint and pre-determination is often referred to as "structure." Some researchers have just a general list of topics that they would like to discuss but let the conversation flow in as natural a way as possible. Others have a list of questions that they ask in as close to the same way as possible with every participant.

Open-ended questions or prompts allow the participant the freedom to respond in many different ways; an example of such a prompt is, “Please tell me about your teaching.” This form of questioning may lead to unexpected information and rich descriptions. However, it may be inefficient in gathering the particular type of data desired by the researcher. At the other end of the spectrum, very structured questions may constrain the responses of the participant to such a degree that the information obtained may be no richer than what could be obtained with a survey. For example, the question, “What textbook do you use?” requires only a couple of words from the respondent. Semi-structured interviews fall in between these two extremes or use questions of both types. An example of a semi-structured prompt might be, “Please describe your most effective physics lesson.” This type of question facilitates comparison across multiple participants but also allows the participant to answer with more depth. One may also choose to ask a structured question and follow it up with a semi-structured question (e.g., “What textbook do you use?” could be followed by “And what do you find to be the strengths of that textbook?”).

Think-aloud vs *Stimulated-recall* vs *Artifact-based:* In a *think-aloud* interview, participants are engaged in an activity (e.g., using a computer simulation, solving a physics problem) and are asked to articulate their thoughts while engaged in the activity (Lewis and Rieman, 1993). This format has been used to understand problem-solving practices, evaluate software, and understand what students are thinking as they interact with computer simulations or physics problems (De Groot, 1965; and Newell and Simon, 1972). In *stimulated recall* interviews, participants watch themselves on video and talk about what they were thinking about the time. This sort of method could be used, for example, with video recordings of teachers teaching to better understand instructional decisions or the knowledge they draw on while teaching (Smith and Neale, 1989). In an *artifact-based* interview, the participant discusses and responds to questions based on a particular artifact, such as a completed homework assignment, a lesson plan, a test, or a graph (Henderson *et al.*, 2007).

25.4.2 Developing an interview protocol

An interview protocol is a written tool that an interviewer uses to guide the interview. It is particularly important to develop an interview protocol when conducting research as a team. When multiple people are separately conducting interviews, a standard protocol that is used by all interviewers will facilitate comparisons across interviews. Generally, an interview protocol includes an introduction, a set of questions or topics, and a closing statement. The introduction should introduce the researcher and the study, describe the interview procedures, and request participants’ permission to be interviewed and, if recording is planned, to be recorded. Writing out an introductory statement ensures that all interview subjects are given the same instructions. (We are omitting here discussion of Institutional Review Board [IRB] requirements, which vary among institutions.)

The second part of the interview protocol (the questions to be asked) is the most important part of the protocol. Careful consideration of the questions will increase the likelihood of useful responses. Often, it is helpful to begin the interview with easier, more structured questions to give the interviewee time to become comfortable before expecting the interviewee to answer more open-ended questions. Finally, the interview protocol should include a closing. This may be as simple as the statement, "Thank you for participating in this interview." However, the closing statement may include a final question asking if the interviewee has anything to add; this can often generate unanticipated yet useful data. One may also request permission to contact the interviewee in the future if additional questions arise.

25.4.3 Conducting and recording the interview

During the interview, the interviewer should be aware of the flow of the conversation and probe the interviewee for clarification or more depth. In semi-structured interviews, the researcher should be flexible in following the protocol, making sure to acquire the desired data but following interesting topics when presented. It is also important for the researcher to be aware that finishing the other person's sentences or asking leading questions are (usually) undesirable in an interview and may significantly bias the data. (Exceptions are discussed below.) It is very natural to impose one's own interests upon an interview and researchers can unintentionally move the interview in a different direction than the interviewee was moving. It is critical to check oneself constantly throughout the interview to make sure that one's comments are kept to a minimum and are as generic as possible. Examples of comments that can lead the interviewee to expand on his or her thinking include generic statements such as a long drawn out "sooo" or explicit statements such as "please say more about that." In all interview situations, the researcher must gain the trust of the participant(s) and establish rapport.

During an interview, it is useful to have paper and pens available in case the participant can express something better in a drawing than in words. A participant may want to draw a graph to describe how students represented their ideas (if the interviewee is a teacher) or to draw diagrams to represent his or her own thinking. These writings can be captured on high-resolution video and/or the researcher can take a high-resolution photo of the writing.

Most practitioners record their interviews. While some participants may find the equipment intrusive, this discomfort generally fades quickly and the quality of data collectable through audio and video recording is far superior to what can be accomplished through note-taking alone. Video recording has the advantage of recording the participant's facial expressions, gestures, and drawings that the participants may create to illustrate their points. Video recording interviews are also important when using think-aloud interviews to investigate how learners interact with computer simulations, physical equipment, or other tools in a learning context.

25.4.4 Purposes of interviews, with examples from the literature

In this section, we describe the different purposes of interviews in PER, and how the specific purpose can guide the methods employed. The amount of prompting and the directness of the prompting that occurs in an interview vary according to the practices and goals of the researchers. At one extreme, the only question asked of the interviewees might be, "Please explain your thinking to me regarding 'topic X.'" More often, interviewers will ask clarifying questions with varying degrees of explicitness. This type of interview is often broadly called a "clinical" interview. At the other extreme, the prompts are frequent and direct and increasingly explicit, in a manner deliberately designed to guide student thinking along certain specific pathways. The latter type has been called a "teaching interview." Teaching interviews may be contrasted to clinical interviews in which interviewers usually avoid, as much as possible, prompting of the interviewee that is designed to *change* his or her initial ideas during the interview. In contrast, in teaching interviews, questions are asked so that interviewees are prompted to think in a certain specific way.

Below, we describe how interviews are used for four different purposes: Assessment validation, understanding students' problem solving, exploring students' and faculty members' experiences, and teaching.

1. Assessment validation (likely to use think-aloud protocol). The purpose of the assessment validation interview is to determine the extent to which the assessment instrument measures what it is intended to measure. There are many different types of validation interviews. Most commonly, interviews are employed to test and validate items on a written survey or diagnostic instrument.

Halloun and Hestenes (1985b) described an interview method used to validate their mechanics conceptual diagnostic test, which after later revisions became the widely used Force Concept Inventory. In these interviews, students were first asked to justify their answers and opinions on the test items related to motion and force; the interviewer would then repeatedly introduce contrary information, asking students for comparisons between different physical situations in an effort to test the stability of students' beliefs. In some cases, typical classroom demonstrations were given of the physical situations described in some of the test items. The interviews yielded evidence of a wide range of student ideas about motion, and the stability of the students' answers bolstered the authors' confidence in the reliability of the test (Halloun and Hestenes, 1985a).

Redish *et al.* (1998) carried out more than 100 h of videotaped student interviews in order to validate their Maryland Physics Expectations Survey (MPEX), which probes students' attitudes, beliefs, and assumptions about physics. The purpose was to confirm that the authors' interpretation of the items on their survey matched the way they were read and interpreted by the students. They asked students (either individually or in groups of two or three) to describe their interpretations of the survey statements and to indicate why they responded in the way that they did, and also asked them to give specific examples to justify their responses. The authors note that students were not always consistent with their responses to what appeared—to the authors—to be similar questions and situations, thus underlining the critical importance of the interview in the validation process.

The creators of the CLASS (Colorado Learning Attitudes about Science Survey), a survey of student attitudes, describe the efforts they took to include a diverse group of students in the interview sample they used to help in validating their survey (Adams *et al.*, 2006). In particular, they took care to select interview subjects from introductory courses catering to the full range of majors, including equal numbers of men and women and 20% non-Caucasian students. The students were first asked to take the survey with pencil and paper. The interviewer then read the survey statements, asking the students to respond and to talk about whatever thoughts each statement elicited. Statements that were unclear or misinterpreted were revised or removed, while expected student ideas were incorporated in a revised version of the survey. A smaller set of follow-up interviews led to a few additional revisions. The interviews provided some new insights into student thinking about physics that had never been previously reported.

Dancy and Beichner (2006) developed an “animated” version of the Force Concept Inventory, in which short and simple computer animation replaced the static diagrams of each question in the original version of the test. During a series of interviews, students were initially shown a random mix of questions, half of which were the original versions and half were the animated versions. They were asked to verbalize their thoughts as they attempted to answer each question while the interviewer remained quiet. The interview data yielded several interesting insights, including (i) students would misread static problems much more frequently than the animated versions, (ii) even when students correctly read a static question, they were often unable to correctly interpret the physical situation—more often than on the animated version—and thus their incorrect answers in these cases were not reflecting the students’ thinking on the question actually intended by the test designers; (iii) the animated versions of the questions were less vague, since they carried more information, and students were less likely to answer such questions with memorized responses.

2. Problem solving (likely to use think-aloud protocol) Interview subjects—usually, but not always, students—are presented with a series of physics tasks, questions, or problems while the interviewer asks them to go through their answers, explaining their thinking in detail as they discuss each of the problems. The interviewer will occasionally ask the subjects for additional clarifying details and may choose to ask specific pointed questions to ensure that important or ambiguous ideas are clearly enunciated by the subject or—if such be the case—that the subjects explicitly express their uncertainty or confusion about the idea. Although the interviewer strives, in general, to minimize the risk of “leading” the subjects to express particular ideas that they may not actually hold, at the same time, it is important to clarify the subjects’ thinking on specific, important issues. In these cases, the purpose of the interviews is not merely to determine whether or not the students can correctly complete the tasks or answer the questions but also to explore nuances of the students’ reasoning process that may not be evident even in open-ended “free” responses on written instruments.

Some of the very early examples of this type of interview in PER may be found in the publications of Reif and McDermott, together with their students and collaborators. For example, Reif *et al.* (1976) provide one of the earliest descriptions of a research interview in PER:

> “To analyze the task of teaching simple problem-solving skills, we began by observing in detail how an individual student goes about a problem-solving task. To do this, we gave the student a

> problem and asked him to solve it while talking aloud about what he was doing. The resulting detailed record, consisting of the student's written solution and his verbal comments (recorded on tape and afterwards transcribed on paper), then constitutes a 'protocol' which one can examine together with any retrospective comments made by the student about his solution process." (Reif *et al.*, 1976, p. 216)

Trowbridge and McDermott (1980) carried out more than 300 "individual demonstration interviews," which, they said, resembled the "clinical interview" pioneered by psychologist Jean Piaget. In this study, such interviews constituted their primary data source. Students were shown multiple trials of metal balls rolling down two separate tracks, one or both of which were inclined, and asked to determine whether or not the balls ever had the same speed. Trowbridge and McDermott described this method as follows:

> "In the individual demonstration interview, the student is confronted with a simple physical situation and asked to respond to a specified sequence of questions. Only simple equipment is used… While the questioning follows a regular format, it allows for exploring any particular aspect of the student's thinking that may be of interest. Each interview lasts from 20 to 30 min and is audiotaped or occasionally videotaped. The dialog is transcribed and analyzed in detail." (Trowbridge and McDermott, 1980, p. 1021)

In a similar fashion, Goldberg and McDermott (1987) carried out interview tasks in which students were shown an optical bench containing an illuminated object, a converging lens, and a screen on which an image of the object was visible. Among the tasks was one in which the screen was removed and the students were then guided to position themselves two meters beyond the initial position of the screen, so they could observe the aerial image visible along the lens axis; students were then asked where that image was located.

3. Exploring students and faculty members' experiences [likely to use a list of questions (protocol)] Interviews to investigate participants' experiences on a wide variety of topics have been used in PER, utilizing both direct and—occasionally—indirect questions to probe subjects' thoughts, reflections, and insights.

For example, Irving and Sayre (2015) carried out a series of interviews focusing on students' perceptions of what it means to be a physicist. The interview sample comprised students recruited from upper-level physics courses; some were re-interviewed more than one year later. Questions asked were related to students' experiences in their current physics classes, their attitudes in physics, future career plans, and finally to a discussion on physicists. (Additional details of this work are discussed later in this chapter.)

Hamerski *et al.* (2022) carried out observations in a high school Advanced Placement physics class in which computational activities were a major focus; in addition, six of the students in the class were interviewed. Interview questions were aimed to elicit feelings about the physics class and computational activities, and to promote discussion of these feelings. The methods used by these authors to code and analyze their interview data are described in the "Coding" section of this chapter.

25.5 TEACHING INTERVIEWS

Corpuz and Rebello (2011) describe the teaching interview as "a mock instructional setting in which the teacher-researcher influences the knowledge construction process of students by providing pedagogically appropriate scaffolding" through a questioning process, sometimes using learning materials or by engaging with multiple students simultaneously. They say that "the goal of the teaching interview is to investigate the variations in the trajectories of student learning and the factors that influence these trajectories." Threats to validity are minimized by emphasizing to the student that there is no intention to provide them a scientifically correct understanding of the phenomena, but instead to explore how they think about phenomena and to probe how they might respond to certain questions. The questions are phrased and sequenced in such a way that they increasingly become leading questions. (Meltzer, 2005, has described analogous techniques developed in other fields during the 1980s, collectively known as "dynamic assessment.")

For example, to probe students' thinking about friction in a *clinical* interview, Corpuz and Rebello asked students to feel both smooth and rough surfaces and sketch them at increasingly small length scales. In the teaching interview, by contrast, students were deliberately asked to sketch the surfaces at the *atomic* level, thus explicitly revealing any student ideas in the context of friction related to atoms.

Corpuz and Rebello have argued that

> "…the teaching interview can serve as a useful step in the design of curriculum materials. By elucidating the fine-grained detail of students' knowledge construction processes, the teaching interview may enable the researchers to create appropriate scaffolding activities that can facilitate learning along a desired conceptual trajectory." (Corpuz and Rebello, 2011, 020103–020107)

25.6 SURVEYS, ARTIFACTS, AND ELECTRONIC SOURCES

25.6.1 Surveys

Surveys are an efficient method for gathering data; since a survey or questionnaire is a method for asking questions of participants, it serves much the same role as an interview. The advantage of surveys over interviews is that they require less time to administer and thus data may be collected from a greater number of participants. Surveys should be tested with a sample from the same population as the participants (often in a think-aloud format) so that researchers can reasonably expect that the questions are interpreted by the participants as the researchers intended. Surveys can be administered in person, through the mail, in paper-and-pencil format, or electronically using websites that are designed to distribute surveys.

Survey formats can vary widely in the nature of the questions asked since responses to questions or statements may be solicited in an open-response "essay" format or instead in "agree-disagree" format using a five- or seven-option scale that ranges from "strongly agree" to "strongly disagree." (The latter format is often referred to as a form of "Likert scale.") Hybrid versions are also possible. Numerous surveys have been developed and validated in PER for a variety of purposes, typically using the "agreement – disagreement" format for efficient data collection; their design principles and development process are described in the original papers written by the authors. Among the most commonly used surveys have been the Maryland Physics Expectations Survey (MPEX) (Redish *et al.*, 1998), the Colorado Learning Attitudes about Science Survey (CLASS) (Adams *et al.*, 2006), and the Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS) (Zwickl *et al.*, 2014). Information and references to numerous other PER assessment survey instruments are available at the PhysPort website (PhysPort, 2022). Some of the design details of MPEX and CLASS are described in the Interview/Assessment Validation section above.

25.6.2 Artifacts

Artifacts can be important sources of data. Artifacts may include, for example, lesson plans, student drawings of their ideas, worked out problem sets, course syllabi, instructor comments on student work, or student or instructor journals. These kinds of artifacts may be collected directly or electronically through scanning or by taking photographs of the artifacts with a digital camera. Photographs of classrooms (including the things that teachers choose to hang on the wall or the layout of the classroom) may provide valuable information about the learning environment. Artifacts such as completed homework assignments and tests can be valuable for checking the validity of claims made from other forms of data.

An example of the use of artifacts is described by Yerushalmi *et al.* (2010). These researchers employed a series of interviews in which physics instructors were asked to compare a series of concrete instructional artifacts, similar to those they were likely to encounter in their teaching environment, and make judgments about them. The artifacts consisted of problem statements, instructor solutions, and student solutions. The instructors were asked to examine several variations of a mechanics problem and to discuss the ways in which the problems were similar to or different from problems they might use in their own teaching on exams, homework or lectures. The authors used the interview data to generate a list of the instructors' teaching and learning goals, along with the value the instructors assigned to the various problem features in the service of their teaching goals.

In a study of the impacts of undergraduate student research experiences, Werth *et al.* (2022) employed a variety of data sources that included responses from weekly post-lab reflection questions that had been assigned to the students. These responses were coded using a thematic coding scheme to find common trends in student responses. They used an *a priori* codebook containing seven codes: affect, authenticity, coding, community, identity, learning, and teamwork, which reflected the authors' motivations for the course. (See Coding section below for discussion of *a priori* coding.) Specific quotes taken from the reflections are provided verbatim in the report, based on their exemplary nature to highlight trends seen within these themes.

25.6.3 Electronic sources of data

Conversations on threaded discussion boards and in electronic chat rooms, as well as students' responses to online homework, are an additional source of data. As online forums become a more prevalent learning context, understanding what happens in these contexts will be important. In addition, the interactions between participants in collaborative online forums may provide important insights into more general questions about teaching, learning, and interaction.

The pandemic era has introduced—or popularized—a number of other electronic data collection modalities. For example, interviews done via Zoom became common, and these conveniently allowed recording through Zoom and enabled automatic transcriptions and the use of closed captions. At the height of the pandemic, researchers even had study participants (student teachers) film their own remote instruction, which could be shared with the researchers during Zoom interviews. Artifacts could be collected through shared whiteboards or shared documents, via Zoom, Google, or other platforms.

Other technical means. Any number of relatively new technologies have been put to use in qualitative research. For example, Franklin and Hermsen (2014) describe the use of "key-capture" technology in the analysis of students' writing and revision processes. Key-capture analysis reveals revisions on the smallest scales, as words are typed, erased, and rewritten.

25.7 OBSERVATIONS

Observations provide an opportunity for a researcher to collect information about an activity as it occurs. During an observation, the following steps must be taken prior to and during a qualitative observation: (1) determining the role that the researcher will play in the learning context, (2) creating or adopting an observation protocol, and (3) recording the observation. Each of these steps is described below.

1. *Determining the role of the researcher.* When observing, the researcher should be aware of his or her role, which can range from fully participating in the context (often termed *participant-observer*) to participating in the learning context as little as possible or not at all. In educational research, adopting the role of participant-observer may mean that the researcher is also a teacher, student, or assistant in the classroom (Spradley, 1980). Researchers may instead choose not to participate in the research context, but rather to remain as unobtrusive as possible. The role of *the non-participant observer* allows the researcher to collect data at the time of the observation. Many researchers take a role that falls somewhere between these two extremes of participant observer and non-participant, or they move flexibly between these roles.
2. *Developing the Observation Protocol.* Observation protocols are developed prior to an observation to focus the observer's attention on relevant aspects of the research context. The structured focus

afforded by observation protocols is especially important in complex learning environments such as classrooms. In addition, for research groups, protocols increase the likelihood that multiple observers will pay attention to the same elements of the context. Researchers may choose to develop an observation protocol or may use a protocol developed by another research team, e.g., the Reformed Teaching Observation Protocol (RTOP) (Sawada *et al.*, 2002). Within each of RTOP's five categories, a handful of specific items are presented and the observer uses a Likert scale to note whether the item is present in the instructional episode or not. This type of protocol quantifies qualitative data. Borrowing protocols facilitates comparisons across multiple research studies but may not match the goals of any one research program as ideally as a new protocol developed for a specific study. However, many widely used existing protocols have undergone extensive validity studies.

3. *Recording observations: field notes and video recording.* One method for recording observations and thoughts is to take field notes, that is, contemporaneous paper or electronic records made by an observer (Creswell, 2005). The current widespread and easy availability of video recording methods has created an environment in which relatively few research teams choose to avoid the use of video, and instead most use some hybrid version of field notes combined with video recording.

Video recording is an efficient method for capturing interactions and actions. As such, video recording of observations is the norm in both the PER community and within the wider education community. Video recordings have many advantages over field notes alone. Field notes are inherently selective because by their very nature, the researchers choose what to write down, documenting what interests them at the time of the observation and failing to document most of what goes on in a classroom. Videos may be watched over and over again by multiple observers, allowing for patterns to be recognized that may have been missed during an observation. In addition, video data provide a richer representation of an event—recording not only talk but also gestures, body language and facial expressions.

Developing a protocol—determining what will be recorded. The video collection is not immune to the effect of the researcher's subjectivity. Video recordings are limited by where the camera is pointed and the field of view. Choosing to follow the teacher with the camera results in the loss of information about what is being done by the students who do not happen to be interacting with the teacher. Zooming in on a student's paper may allow the researcher to read what a student writes, but may miss gestures that a student uses to explain his or her reasoning to a classmate (Erickson, 2006; and Jacobs *et al.*, 2007). A second camera might then be used to capture the entire classroom with all participants in view.

One should also consider that the presence of a video camera may impact the participants' behaviors. Jordan and Henderson (1995) suggest that when the camera does not have an operator behind it, participants get accustomed to it more quickly, treating it as any other piece of furniture. Not having an operator behind the camera, however, limits control over where the camera points.

One common problem with video data collection is equipment failure, most commonly inadequate audio quality. The microphone(s) may not be adequate for recording multiple overlapping voices or for the range of distances often necessary in classrooms. Extraneous noises (e.g., papers rustling) near the microphone may be inaudible to an observer but may be so loud on the recording that the participants' voices are unintelligible. Whether or not voices can be heard properly can be checked with headphones that are plugged into the recording device. A wide variety of microphone types are available, and proper selection and testing are essential for successful data collection.

An example of a combination of field notes with video recording is described by Daane *et al.* (2015). Video recordings were made of teachers working in small groups in a professional development course. Video episodes in which learners were engaged with the phenomenon of interest (energy dissipation) were identified through (i) initial field observations recorded by the videographers, and (ii) a search for key terms in the field notes that could relate to the topic of interest. Several researchers studied facial expressions, interactions between participants, bodily behavior, and other indicators. The reactions from the learners were indicated by changes in various participants' verbal and behavioral interactions with one another. Through a detailed analysis of the video transcripts, including both the spoken words and the participants' gestures (e.g., making a "fanning" motion to represent production of thermal energy, and a squatting motion to indicate energy dissipation by a person), the authors supported their claim that some learners expect that energy associated with a perceptible indicator will be associated with another perceptible indicator when the energy transforms into another form. The analytic framework adopted by the researchers was that the general properties of an event or phenomenon emerge from the specifics of a particular case, rather than from the patterns that emerge across cases. Therefore, while numerous instances were identified from the video transcripts to provide evidence for their primary claim, there was no explicit attempt to identify a general "pattern" across cases through, for example, relative frequency counts of specific observations in different cases.

25.8 PROCESSING DATA

Once the data have been collected (or during collection), a qualitative researcher begins the process of condensing or summarizing the data. In this section, we discuss the first steps towards condensing the raw data.

Video and audio recordings are generally processed into a text-based form, which can be further analyzed. Processing data not only prepares the data to be analyzed but also is the first step of analysis. Erickson (2006) suggests that the first stage of video analysis should be reviewing the video recording continuously without stopping it and writing field notes known as "content logs." A content log is a list of events that are on the tape. The level of detail may vary depending on the needs of the researcher (for an example, see Jordan and Henderson, 1995).

In order to analyze effectively, video data is often turned into a typed transcript of all the words that the study subjects said and the gestures they made. A transcript of video or audio data represents what was recorded, and while researchers may return to the video frequently during analysis, much of the analysis is done on and with typed transcripts. Thus, deciding what and how to transcribe should be done with care, and the actual transcript *of the specific sections chosen for transcription* should be as complete and accurate as possible.

Depending on the research goals, one may choose to transcribe all the data or only relevant sections of video or audio recordings. As stated earlier, one way that researchers know that they have enough data is that additional data do not reveal anything new. The selection of what to transcribe should be done purposefully and carefully, if at all. One may limit the data transcribed to a particular activity or lesson or by interesting interactions. One should select events to transcribe cautiously, especially if the decisions involve medium-to-high inference. For example, a researcher who chooses to only transcribe events in which students are engaged in "sense-making" has already made some decisions, often based on tacit theoretical perspectives, about what constitutes sense-making behavior, introducing bias into the data set by making decisions about what to transcribe.

Even if one chooses to transcribe *all* the data collected, decisions about what to transcribe must be made. One may choose to transcribe just the words or to transcribe the words *and* gestures, or even the words, gestures and direction of gaze of the participants (Scherr, 2008). One must also make decisions about whether to pay attention to changes in the tone of voice or speed of talk. Because of the myriad decisions made in choosing what to transcribe and what not to transcribe, transcribing is the first step of analysis; it is not just pre-processing. When deciding the level of detail to transcribe, a researcher must decide what will count as data and what will not.

Software for transcribing. A variety of software packages have been developed to aid in transcription, and voice recognition software is now widely available and reasonably accurate. (However, electronically generated transcripts generally still require careful editing.) In considering which to use, one should consider a number of factors including cost (which ranges from free to thousands of dollars), the format of the data, whether the transcript will need to be turned into subtitles for presentation, and whether one wants to be able to analyze and transcribe with the same data.

25.9 CODING AND ANALYZING DATA

Ultimately, the goal of conducting qualitative research is to find meaning. This is often done through the process of coding text-based data in order to categorize the data and to find meaning within it. In this section, we focus on the mechanics and common methods of coding transcripts. However, one should keep in mind that finding meaning in qualitative data may take a variety of other forms.

While some of the summarizing and condensing of data occurs during the processing stage, the majority occurs during the coding and analysis stages. Note that in qualitative analysis, analysis does not generally occur in discrete steps, but rather the many strategies discussed occur somewhat simultaneously.

Before beginning, it is worthwhile mentioning the critique of Hammer and Berland (2014), who argue that researchers should not present the results of coding as if they are data; instead, results of coding should be seen as *claims* about the data. The implication of this argument—following Schoenfeld (1992)—is that researchers should provide sufficient detail about their coding scheme, with adequate samples of their data, to enable readers who wish to do so to apply the coding method themselves to confirm the authors' analysis. Hammer and Berland also advocate a discussion of borderline cases and of the uncertainties implied by measures of interrater reliability.

25.9.1 Coding overview

Coding transcripts or other text-based data is the process of going through a transcript in detail in hopes of finding words, statements or events that can be sorted and labeled using a cover term (code). Ultimately, the researcher will use these codes to find patterns and meaning in the data (Creswell, 2005).

Codes are labels and may pertain to many different categories. Codes could, for example, relate to specific concepts, activities, or ideas. The size of the text segment that is assigned a code may be a word, a phrase, a sentence, or a "turn of speech," that is, all that an individual says while it is their turn to speak following another speaker, and before the next speaker begins (Lerner, 2004). A code could cover multiple turns of speech by an individual speaker, or alternatively, multiple codes could correspond to a single turn of speech.

25.9.2 Developing a coding scheme

A priori vs generative codes. Depending on the research question, coding schemes are developed either before coding (*a priori)* or during coding (generative). *A priori* codes are useful when the researcher is looking for something in particular in the data, or is testing hypotheses. Generative codes are useful when the researcher is looking at the data to discover what is there. In actuality, coding schemes are often a combination of *a priori* and generative codes.

A priori coding schemes may be developed out of prior work by other scholars, one's own theories and assumptions, or some other pre-determined schema. The researchers might, for example, start with pre-determined sets of common student ideas from the literature, look through their data, and note places where they saw evidence for one of these specific ideas. Because *a priori* codes are determined *prior* to coding, the emergent nature of qualitative research means one must be willing to modify these pre-determined codes if the data do not support them. In contrast, *generative codes* are generated as part of the coding process. Both generative and *a priori* methods are useful for different research projects. What is important is that the researchers are clear about what they are using and that this is reported in their methods.

Generative coding. Often in qualitative research, the goal is to understand a particular phenomenon by looking at data and exploring what is there. Generative coding develops through the process of examining and analyzing data; it is iterative, and specific methods of developing codes vary. For example, initial examination of the data may yield several preliminary codes. As the researcher becomes more and more familiar with his or her data, new categories may emerge or multiple categories may collapse. Several examples of this process drawn from the literature are discussed below. Computer software designed for the purpose of coding allows researchers to engage in a similar process electronically; many such software packages are currently available, including some that are available free of charge.

25.9.3 Working with codes

After coding, the qualitative researcher attempts to establish links between the codes and to further reduce the codes with the goal of obtaining the minimum number of independent categories to represent the data. Ultimately, the links, codes, and categories should lead to a description or explanation of the way that the study participants think, talk, or behave. The smaller number of categories is more manageable to work with. Once codes have been created and categorized, one begins to look for larger patterns in and across the codes. There are many methods for looking at and representing codes and categories to make trends more visible; among the more common are *taxonomic domains* and *componential analysis*.

Taxonomic Domains. Codes that are hierarchical in nature can be organized into a taxonomy. In a taxonomic domain, each domain contains a cover term (e.g., "participants") and a number of subterms (e.g., "students," "faculty," "other") such that a taxonomy can be represented in a way that shows the hierarchical relationship between the codes. Bloom's taxonomy is an example of a taxonomic domain (Bloom, 1956).

Componential Analysis. In *componential analysis*, one compares the presence or absence of selected attributes, for example, the specific terms that participants use when referring to learning activities (e.g., lecture, lab, homework) or the type of talk (e.g., metacognitive, sense-making, off-topic) that is present during different types of learning activities; for a full description see Spradley (1980). The goal of this type of analysis is to distinguish items of interest.

The fact that there is a very broad range of appropriate methodological approaches in qualitative data analysis means that a researcher must be very explicit and detailed in describing the methods that lead to the researcher's claims. It is not enough to state that the data were analyzed through "qualitative methods" because this does not sufficiently explain the process. This point is emphasized by Hammer and Berland (2014), as discussed above. For example, van Zee *et al.* (2005) make the qualitative method they used very apparent. They discuss how they divided the data into episodes and provide examples of their categorization. A number of other examples from the literature are discussed below.

The codes and patterns observed in the codes can be used to develop a *descriptive* model of the phenomenon being investigated that does not necessarily include quantitative measures or attempt to explain the observed phenomena. Alternatively, a researcher may use the frequencies of the codes themselves to make descriptive statements; examples are given below. Sometimes, a descriptive model is the researcher's desired final outcome. The utility of such research is exemplified by Walsh *et al.* (2007), who provides a descriptive model of students' approaches to problem solving and compare it to a similar descriptive model developed by Tuminaro and Redish (2007). However, it is often the case that the researcher seeks to *explain* the phenomenon involving human subjects.

In order to create an explanatory model, researchers must represent and re-represent their findings and link these findings to one another and to a learning theory or "explanatory framework." To further this goal, after developing a coding scheme, it may be helpful to represent the data in a concise summary form, for instance, using tables, graphs, or diagrams. Such representations might be critical in helping the researcher make inferences from the data and in ultimately constructing an explanatory model, when appropriate. While many qualitative research designs do not rely on counting anything, one may often want to integrate quantities into qualitative designs. One may count the times that students are engaged in a particular activity or count the number of instances of students supporting claims with evidence (Chi, 1997).

As is the case in any kind of research, in qualitative study claims are made and must be supported with evidence from the data. Claims in qualitative research are often based on a preponderance of similar occurrences in the data. These claims may be supported by frequency measures. For example, qualitative researchers might interview 40 physics students and make the claim that these students were thinking about electric current as a fluid. The researchers might even provide some illustrative examples of what students said that led them to believe that they were thinking about electricity as a fluid. Frequency measures may help the researchers' audience understand what they mean when they say that "the students" were thinking about electricity as a fluid. They could state, for example, that 30 of the 40 students interviewed, or 75% of the students interviewed, seemed to view electricity in this way.

While numerical values are useful in supporting inferences made from the data, it is not always necessary to use numerical data to support claims or inferences. For example, Tuminaro and Redish (2007) claim to have found five different "epistemic games" that students play while solving physics problems in a small group setting. One example of an epistemic game is "recursive plug and chug." Tuminaro and Redish used a careful analysis of transcript data and thoughtful use of transcript excerpts to make the case for the existence of certain characteristics of the "recursive plug and chug" game and its distinction from other epistemic games. While quantities can be extremely helpful in making claims, they are certainly not necessary for making strong and reliable claims in qualitative research. It should be noted, however, that inferences are often tied to the theoretical perspective that serves as a lens through which the researcher views the data.

A number of more detailed illustrations of coding and analysis are discussed below.

25.9.4 Examples of coding schemes

25.9.4.1 Categorizing interview responses to "fact" questions

Perhaps the most straightforward form of interview coding occurs when students are asked to explain their answers and their reasoning on physics tasks which have clear-cut answers. However, even here, the coding is designed to reflect the nuances of the students' reasoning process, and *not* merely "right" or "wrong" answers. Some of the very early examples of this type of coding in PER may be found in the publications of McDermott with her students and collaborators. For example, Trowbridge and McDermott (1980) carried out individual demonstration interviews in which students were shown multiple trials of metal balls rolling down two separate tracks, one or both of which were inclined, and asked to determine whether or not the balls ever had the same speed. Students' responses were coded according to a three-category scheme in which Category 2 corresponded to giving a correct response on the first or second trial with no confusion of speed and position; Category 1 indicated that, after confusion on initial trials, students were later able to reverse their initial judgment and provide acceptable responses, and Category 0 indicated that students were unable to describe a procedure for determining when the ball speeds were equal, even after three or more trials. Among the outcomes of the work was a finding that failure on the interview tasks was almost invariably due to improper use by the students of a *position* criterion to determine relative velocity (e.g., that balls passing each other had identical velocities). This criterion was employed not because students were mistaking one fully developed concept (position) for another (velocity) but instead because they were making use of undifferentiated "protoconcepts" that included a "repertoire of procedures, vocabulary, associations, and analogies." This latter finding—that students were often employing protoconcepts rather than fully developed concepts—was presumably facilitated by the use of the intermediate Category 1, whereas it may have been less obvious had a simplistic Correct/Incorrect categorization been used instead.

In a similar fashion, Goldberg and McDermott (1987) carried out interview tasks in which students were shown an optical bench containing an illuminated object, a converging lens, and a screen on which an image of the object was visible. Among the tasks was one in which the screen was removed and the students were then guided to position themselves two meters beyond the initial position of the screen, so they could observe the aerial image visible along the lens axis; students were then asked where that image was located. The students' responses covered all possibilities, including (i) between light bulb and lens; (ii) on the lens; (iii) a non-specific position between the lens and eye; and (iv) at the same position as screen [correct]. The coding scheme was *not* determined in advance—although it may have seemed obvious enough to do so—but was generated from the students' responses themselves. Interviewers also noticed that students were unable to move from category (iii) to category (iv) even when repeatedly shown the screen being moved into and out of the light path so that the screen image would re-form. Had the researchers used a pre-determined scheme that excluded either category (iii) or (iv), this important finding might not have been possible.

An example of using interview data without a detailed formal coding procedure is provided by Tu *et al.* (2021) in their study of student difficulties with bound- and scattering-state problems in quantum mechanics. They collected both students' solutions on written exams to identify common student difficulties, and interview data "to gain deeper insight into the nature of these [common] difficulties." In one task, students were asked to determine the possible energies of the bound states for a particle interacting with a one-dimensional well potential. During interviews, students were asked to formulate and articulate their thought processes, and were asked to provide justifications for the various specific calculations they performed as they worked through the problem. About half of the students were able to correctly state that a bound state could only occur if the particle energy was less than the potential energy at infinity, but only a few students could state that no physically acceptable solution would exist if the energy was less than the minimum value of the potential. Many students were unable to clearly distinguish the conditions for bound and scattering states. Although responses on the written version of the problem reflected the proportions of students who either obtained solutions or made various errors, those written responses could not provide the insights into student thinking regarding bound and scattering states that were obtained through the interviews.

25.9.4.2 Categorizing interview responses according to the *a priori* scheme

Brookes and Etkina (2009) devised a grammatical scheme for classifying language used to define the properties of force, based on historical analysis of the development of the force concept; this scheme classified statements about the nature of force as identifying the "location" of force as either internal or external, and the "role" of force as either active or passive. (A particular description might then be categorized as, for example, "active; external." Students' use of the prepositions "on" or "to" would suggest that the location of the force is external, while saying that an object is "doing something" to another object would suggest that the role of the force is active.) They then used this scheme to classify excerpts from interviews extracted from 12 previously published papers or books in which students' reasoning about force and motion problems was reported. The authors thus coded 49 student explanations, finding 33% categorizing the role of force as "active," 47% as "passive," and 20% as ambiguous. Similarly, they coded 6% of the locations of force as "external," 27% as "internal," and 67% as "ambiguous." (In this context, "internal" means that the students interpreted the force as being *inside* the object or a *property* of the object.) The instructional implication drawn by the researchers was that many students who manifested a conception of force as a passive participant in events, rather than as an active mover, may therefore conceive of force as a *property* of motion, in contrast to the Newtonian view of force as an interaction that can *alter* the motion. In this case, the use of an *a priori* coding scheme allowed researchers to uncover potentially important aspects of students' reasoning about forces.

The same authors (Brookes and Etkina, 2015) carried out 10 student interviews to analyze students' reasoning regarding heat; the authors found patterns in the interview transcripts that were in part consistent with experts' language but that mostly conflicted with it. For example, statements that defined heat as "energy in transit" or "the quantity in the equation" were classified as "operational," consistent with experts' definitions. However, statements defining heat either as a substance on its own or as a form of energy without further elaboration were both classified as a "caloric/form of energy" definition. Here, although the "caloric" coding category was itself drawn from the literature, the specific criteria for including a particular statement in that category were determined through analysis of the interview transcripts.

In a study reported by Gupta *et al.* (2014), the authors go into some detail about the methods they used to analyze language in transcripts of student interactions, including "predicate analysis" and analysis of grammatical metaphors. Predicate analysis explores the predicates (attributes, relations, or properties) associated with certain terms such as force, heat, and light to infer the "ontological category" (the specific nature or type) identified with that term. They cite an example given by Slotta and Chi (2006): if a student should describe current as something that can move and be consumed, that implies the identification of the current as a "substance-like" entity rather than a process-like entity. Similarly, grammatical structures may be used for a similar purpose. For example, Gupta *et al*. note that a student's statement that "gravity's still pulling the heavier thing down" implies the identification of gravity as an agent or an active participant—a "matter" ontology.

25.9.4.3 Examples of generative coding

An example of a coding scheme that was based on categories that were generated solely through analysis of the initial data is provided by Pulgar *et al.* (2021), who studied four groups of college physics students engaged in a collaborative task to create physics problems appropriate for high school students. The researchers segmented transcripts of recorded group discussions by separating speech into individual clauses (a clause typically contains a subject plus a verb phrase), and identifying the topic associated with each clause. Their initial coding scheme was based on analysis of 25% of the recorded data and contained 12 categories, which were then reviewed and reduced to only six categories, including "physics concepts and procedures," "discussing magnitudes and units" and "algebraic procedures." Once the initial coding scheme had been developed, an additional researcher was brought in, guided in the use of the scheme, and additional data were coded. The relative frequency with which each category was identified in the coded data was determined, for example, physics concepts and procedures, 34.9%; the problem context and wording, 21.4%. These frequencies were also determined separately for each of the four student groups, and displayed in a bar chart along with the overall averages, thus indicating both the average discussion time for each topic and the intergroup variation in those times. Through this analysis, the authors were able to provide a sense of the relative time devoted to various physics and task-related ideas utilized when composing physics problems for high school students, as well as the degree of consistency of that time from one group to another. It seems unlikely that an *a priori* coding scheme could have been as effective in accurately capturing the various aspects of students' thinking in this case, since there was little or no theory or previous research that could have guided the development of an adequate scheme.

Hamerski *et al.* (2022) carried out observations in a high school Advanced Placement physics class in which computational activities were a major focus; in addition, six of the students in the class were interviewed. Interview questions were aimed to elicit feelings about physics class and computational activities, and to promote discussion of these feelings. The researchers identified "episodes" from each of their interview transcripts where the discussion centered around these feelings, tracing out patterns across the different episodes and interviews. Each pattern was named according to the common experience or challenge that it represented for the students. Initially, six categories were identified, one of which was "Stress/Frustration." The authors found that students often saw the stress they felt as uncalled for, in that those students felt they already knew the relevant physics concepts and computation was just forcing them to jump through additional hoops, generating unexpected and frustrating difficulties. The authors provided extensive verbatim quotes from two of the six student interviews to support their findings on this category, noting that other students also experienced stress but articulated it in different terms.

It is useful to relate the analysis method described here to the explicitly stated goals of this study. The authors state that their aim was not to generalize their results "to any sort of population," but instead "to describe the variety of challenges students faced" in this class. The authors justified this aim by noting that their study was perhaps the first to focus specifically on students' perceptions of the integration of computation into STEM classes "and the impacts on their affect," thus situating their study's goals as an initial attempt to fill that knowledge gap. In this way, one could justify the absence of any effort to assess the *frequency* with which certain student feelings occurred or how prevalent those feelings might have been throughout the student sample. Instead, by clearly identifying certain patterns that at least *some* of the interview subjects manifested, the authors arguably succeeded in their goal of describing the "variety of challenges" faced by the students. A study with a different goal—for example, a goal of determining a pattern that could be generalized to certain specific populations—might instead have needed to apply some sort of statistical analysis to their interview findings. One can also see why the coding scheme applied in this case was determined through the data rather than in advance, since the lack of previous research on the topic provided few grounds for guessing what specific challenges were likely to come up during the interviews with the students.

25.10 THEORETICAL PERSPECTIVES AND FRAMEWORKS: STUDY TYPES

A theoretical perspective or framework, in our context, can be defined as a broad, systematic outlook and strategy for addressing the challenges of research on thinking and learning; these perspectives and frameworks motivate and guide a variety of specific methods for carrying out investigations. More often than not, "choosing" a theoretical perspective has more to do with *becoming aware* of one's own prior thoughts and beliefs about how people learn than with actually selecting a theoretical perspective from a list. Our tacit theories are always lurking in the background and have great influence on how we see things and on the claims we make; such tacit theories often tend to evolve for individual researchers as they explore ever more diverse aspects of students' learning through their careers. Under this heading, we also include both *practical* and *conceptual frameworks* as specific varieties of systematic strategies for planning investigations and analyzing data.

Many claims can be made and many trends can be found in almost any data set. However, trends may or may not be interesting or useful. So while many trends can be found in data—and the very same data set may yield multiple different interpretations depending on how it is analyzed—it is up to the researcher to frame the research in a way in which trends in the data have some meaning and lend support to a broader issue of interest to the community. The strategy one adopts for the choices one makes is sometimes identified as theoretical framing. One's theoretical perspective can influence the research *design* and *methods.* In qualitative research, it is difficult to conduct a controlled study mainly because it is usually impossible to create a control group when aspects both intrinsic and extrinsic to the individual or context being studied are considered. Inevitably, choices must be made on which aspects of the learning process to focus. This process of choosing can be guided by theoretical framing.

There have been many approaches to the problem of theoretical framing in qualitative research and educational research more generally, including, among many others, Vygotsky (1962), Rumelhart (2017), Posner *et al.* (1982), Posner and Strike (1992), diSessa (1988); and diSessa and Sherin (1998). Many others could be cited, along with hundreds of studies that applied these various theoretical framings in practical research. Here, we simply provide a few illustrative examples of PER studies that adopted different framings in order to attain their objectives.

Mestre and his colleagues provide an example of research that is framed with the "coordination class" theory, defined by diSessa and Sherin (1998) as "a systematic collection of strategies for reading a certain type of information from the world." Mestre *et al.* applied the coordination class theory to explain why students' answers to isomorphic questions are dependent on the type of representation that is provided (a simulation of two balls rolling down a ramp vs a sequence of static snapshots of the ball's motion). They used this framing to argue that *the different representations cued different cognitive elements for the students* and therefore, the students responded differently in the isomorphic situations (Mestre *et al.*, 2004) (see also Hernandez *et al.*, 2021).

Henderson (2005) and Chasteen and Chattergoon (2020) use Roger's model of the "innovation decision-making process" (Rogers, 2003) as an analytical framework for analyzing data on physics faculty members' instructional change process. Although they modified the theoretical framework slightly, it was useful for analyzing data to guide their understanding of the evolution of faculty members' teaching practices.

Other researchers use practical frameworks or conceptual frameworks to frame their research; below we briefly outline these approaches.

Practical Frameworks are constructed with practical improvements in mind, where findings lead to changes in practice, informed by accumulated practical knowledge. An example of this is the work of the Physics Education Research Group at the University of Washington (Trowbridge and McDermott, 1980; and Heron, 2004). These types of studies attempt to be agnostic about a theoretical stance, and instead seek to investigate what students are thinking regarding a specific concept in physics, given a specific problem statement. As a result of the new understanding gained from the research on students' ideas or difficulties, instructional interventions are designed and their effects measured.

If the measured effects are relatively large, the researchers conclude that the “difficulties” that they identified were treated by the intervention since the intervention which was based on these inferences was successful.

Conceptual Frameworks consist of arguments that include several different theoretical perspectives and/or robust findings from practical research, and a justification for adopting these perspectives or findings for the current study. For example, Strubbe *et al.* (2020) used a conceptual framework bringing together Self-Determination Theory (Ryan and Deci, 2000) with 7 principles of teaching and learning distilled from the book “How Learning Works” (Ambrose *et al.*, 2010) (with an extra principle that the authors added). They used this framework to establish a conceptual framework for faculty agency around teaching. This framework was used to analyze interviews with physics faculty members to reveal the agency with which they approached their instructional decision making. They conclude their study with several recommendations for curriculum developers and those interested in curriculum and course transformation, to draw on faculty members strengths, ideas, and intentions. These recommendations have practical application and are derived from bringing two perspectives together. An advantage of this type of framing is its flexibility and the fact that conceptual frameworks are often based on the nature of the specific data and findings of the study at hand. Other PER publications explicitly discuss the utility of multi-perspective conceptual frameworks, for example, Rebello *et al*. (2005).

25.10.1 Examples of issues addressed by perspectives and frameworks

Among the broad issues that are addressed differently by different theoretical framings in PER are (1) the degree of “stability,” or consistency of application, of various ideas, beliefs, and behaviors about physics manifested by student learners; and (2) the degree to which socio-cultural factors influence and determine physics learning, in contrast to purely cognitive factors related to an individual’s thinking process. Some studies might focus, for example, on specific student physics ideas on the assumption that such ideas are *relatively* stable and general, for example, the idea that the continued motion of an object requires a continuous push or pull on the object, e.g., Halloun and Hestenes (1985b). Other studies might emphasize instead narrower ideas about mechanisms that are more general and context-independent, that may be activated or “cued” *in certain physical contexts* to generate commonly observed difficulties related to force and motion, e.g., diSessa (1993). Yet another approach is to examine the *evolution* of such student ideas during the learning process, such as the study of Thornton (1997) discussed below.

An example of a socio-cultural perspective that has been used in PER is Vygotsky’s theory of concept formation (Vygotsky, 1962). According to this perspective, learners bring experience-based generalizations to the learning process that are expressed in everyday language, and attempt to mesh those generalizations with “academic” concepts that are typically abstracted from *many* particular experiences and expressed in *formal* language. Learning is then seen as the process by which academic concepts become connected to the experiences of the learner, and related experience-based concepts become generalized beyond the experiences to which they are tied (Otero, 2004).

An example of a study within a conceptual framework is that of Thornton (1997), who reported a detailed investigation into students' evolving views on force and motion, probing the views of the same students over the course of a semester. Thornton's framing was to explore students' cognitive processes, but in a way that viewed these processes as undergoing change and development. The data source in this case was "carefully constructed short answer evaluations (multiple-choice questions written in natural language, questions using graphical representations, and short written answers)" where the researchers "look for correlations among the answers of a number of questions. All answers, 'right or wrong,' are designed to be significant." Thornton specifically examined the changes in students' conceptual views over time, something he called "conceptual dynamics." Thornton provided numerous descriptions of student views in the actual words of the students, drawn from their answers to the open-response questions on the assessment tool. Detailed quantitative analysis was also performed on their responses to multiple-choice questions, and the responses to those questions were found to be consistent with the students' explanations of their thinking.

Some years later, the same author (Thornton, 2004) carried out a very different type of study, this time focusing specifically on student interactions and behaviors—a different perspective in the service of the same goal of improving student learning. In this investigation, nine groups of students were recorded on video and audiotape during at least five of their 10 two-hour labs during a one-semester introductory physics course; many hours of videotape had to be reviewed in detail. Individual and group behaviors were studied, and behavior characteristics of individual students were compared to their performance on the FMCE mechanics conceptual test (Thornton and Sokoloff, 1998). Students were assigned to coding groups depending on their performance on the pre- and post-instruction FMCE; students who initially scored "low" on the pretest (below 25% correct) but who, at the end of the course, scored "high" (78% or above) were designated "low-high." Their behaviors were contrasted, in particular, to students in the "low-low" group who both started and finished with scores below 25%, and students in the "med-high" group whose pretest scores were above 25% but below 78%, and who scored "high" on the post-test at the end of the course.

Among the behaviors studied was the frequency with which students asked either "open" or "closed" questions, defined by the authors as follows: A closed question is one that can be answered by a single word or phrase and does not invite exploration, for example, "What is the sign of the acceleration?" An open question or statement cannot be answered by a single word or phrase and invites exploration, for example: "How can a collision result in changed motion since the forces between the two objects are equal and opposite?" Students in the low-low group were found to have asked many more closed questions, and many fewer open questions, than students in the med-high or low-high groups. Similar differences were found in the number of explanations related to cause or principle that students offered; those in the med-high and low-high groups offered many more such explanations, whether *correct/consistent* or *incorrect/not consistent*, than did students in the low-low group. The authors conclude by noting that students who did learn effectively—those in the low-high and med-high groups—were not necessarily "more involved" with the learning activities than those in the low-low group who did not learn; it was the *specific nature of their learning behaviors* that was associated with the differences in learning effectiveness.

We describe below two types of studies that are not representative of distinct theoretical framings *per se* but instead reflect specific orientations toward certain aspects of learning that are seen as particularly important.

Phenomenography. In a form of research known as *phenomenography*, the emphasis is on exploring the different ways in which people (or groups) experience and reflect on various phenomena and activities to which they're exposed, specifically how different people perceive or understand the *same* phenomenon. An example of such a study is Walsh *et al.* (2007), in which a thorough explanation of the analysis methods is provided. Another more recent example is the study by Irving and Sayre (2015), who provides a detailed description of the coding and analysis methods they used in a series of interviews focusing on students' perceptions of what it means to be a physicist. The interview sample was composed of students recruited from upper-level physics courses. Questions asked were related to students' experiences in their current physics classes, their attitudes in physics, future career plans, and finally to a discussion on physicists. The authors describe their work as being based on phenomenographic research methodology, identifying a limited number of qualitatively different and logically interrelated ways in which a phenomenon or situation is experienced or perceived. Pursuing this approach, researchers read each interview transcript multiple times, with each reading focusing on one particular aspect of the transcript or theme. The themes emerged through the reading process, and it was expected that each theme would emerge from multiple places in the interviews. For example, one important and relevant emergent theme was students' conceptions of when they will consider themselves as physicists. Similarities and differences among the students in addressing each theme were explored to identify perceptions that were either shared or contrasting. Separate categories of description were then formed such that each category encompassed the variations in the students' perceptions. The entire analysis process was performed by each of the two researchers, and final category descriptions emerged through a negotiation and review process. The category of perceptions expressed by the largest number of students was that of "conducting research," which could be expressed as "research is very important to being a physicist and when I am conducting research I will be a physicist." The authors stated their belief that phenomenographic research methodology provides the opportunity to explore student and faculty experiences at a deep level and to discover the nuances between their experiences.

Ethnography. In *ethnographic* research, observations and interviews are used to explore the functioning of both individuals and groups; in anthropological research, the groups may be entire societies, while in PER, the focus is narrower. It is sometimes said that ethnographers "observe life as it happens," in contrast to the sometimes artificial and manipulated context of a highly controlled experiment. For example, Brookes *et al.* (2021) describe an ethnographic study of 30 students in an introductory physics course working in groups of three, with two groups to each table. Each of the five tables was recorded using a miniature camera mounted in the ceiling. A voice recorder was placed in the middle of the table to acquire the best possible audio. Groups were coded on, among other things, the proportion of their time spent in conversation with each other to address the problems they had been given. They were also coded on the "effectiveness" of the groups in correctly addressing the key conceptual points of the exercise, with additional weight given if they verbally justified their reasoning. A qualitative analysis of the recordings was performed to explore why groups that spent roughly equal proportions of time in conversation differed widely in their effectiveness. The researchers noted that conversations within the more effective groups showed group members making a relatively high proportion of "hedged" statements, indicating limited certainty or mild disagreement, as well as asking questions that served to drive the conversation. In contrast, members of less effective groups tended to make more statements conveying "firm" or strong claims of fact or disagreement, as well as asking questions indicating helplessness or general confusion. A detailed quantitative analysis yielded findings consistent with the qualitative analysis. [It is interesting to compare the findings of this study with those of Thornton (2004) described above.]

A study by Wu *et al.* (2022) used a coding process for lab-session interview data that utilized BORIS, a behavioral analysis software package that allows for the easy logging of social interactions. Researchers coded the duration and directionality (who initiated each interaction) of each interaction between the instructor and students. Using this coding scheme, the coders were able to watch the video from each lab session at a fast-forwarded rate, recording the beginning, ending, and directionality of all interactions for each roughly two-hour lab session in about an hour. This method is more efficient than many other methods; however, it fails to capture the content of the interactions, providing breadth but not depth of instructor interactions.

For a more detailed discussion on theoretical perspectives, the reader is directed to research review articles (Fosnot and Perry, 2005; Greeno *et al.*, 1996; Scott *et al.*, 2007; and Anderson, 2007) as well as to Otero and Harlow (2009).

25.11 VALIDITY AND RELIABILITY IN QUALITATIVE RESEARCH

25.11.1 Validity

Validity is the trustworthiness of inferences drawn from data. In qualitative research, "internal" validity can be addressed by implementing research methods that increase the likelihood of eliciting an accurate view of the participants' reality. Validity can be increased by using multiple sources of data to support claims and conclusions (triangulation), such as students' written work, classroom video recording, and audio/video recorded interviews.

25.11.2 Reliability

Reliability refers to the extent to which studies can be replicated. The nature of qualitative research limits reliability in the traditional sense. Qualitative PER is multilayered, constantly changing, and involves multiple populations. Internal reliability requires that within a single study, multiple observers of the same phenomenon come to the same conclusion (Howe and Eisenhart, 1990). Instead of seeking replication in the traditional sense, qualitative research often seeks to describe a slice of life as accurately as possible. When a researcher describes a slice of life, she provides a “thick” description of elements in the environment, what the actors say and do, and many other relevant interactions. [See Sec. 25.2 above, as well as (Geertz, 1973).] One way to test for reliability is for the researchers to sit with the research subjects to check the accuracy of their descriptions of the actors, their actions, and the context in which they act.

25.11.3 Gaining confidence in claims

Qualitative researchers use several strategies to gain confidence in their claims. During coding, researchers check for *inter-rater reliability*. During analysis, they *triangulate* data. Following analysis, they engage in *member checking*. These three strategies are described below.

Inter-rater reliability is a process to make sure that the coding process is as objective as possible. Once the coding scheme has been developed and the codes have been described, researchers may ask another person to use the coding scheme to code a section of transcript. The two researchers then compare the degree to which their coding is similar. If there is great disagreement, the coding scheme requires refinement or better definitions of the codes.

Triangulating claims is a process of comparing data of different types (e.g., interview and observation) and across multiple times (comparing one observation to another) and multiple participants. One may look at both observations and interviews to gain a better understanding of a student’s ideas.

Member checking is the process of checking interpretations with participants. A researcher might ask a participant to review research findings and interpretations and provide feedback on whether the findings and descriptions are accurate and complete.

Other strategies for gaining confidence in claims

- Make all assumptions and perspectives explicit to yourself and others.
- Keep good notes to see how your perspective has changed from the initial assumptions.
- Describe your role in the research setting.
- Utilize multiple viewings of video recordings and multiple researchers.

25.12 SUMMARY

Qualitative research is a time-consuming and rewarding process. Like other types of research, it is systematic; claims are made and supported, and measures are taken to gain confidence in one's claims. In all reports of qualitative research it is best to describe not only the research context in great detail but also the theoretical perspective that the researcher thinks might influence the decisions she makes throughout the research process. Like other types of research, one's results and inferences should be transparent to the reader. Although qualitative research is not generalizable or predictive in ways that quantitative research is, we hope to have presented enough information for the reader to see the utility and value of qualitative research.

REFERENCES

Adams, W. K. *et al.*, Phys. Rev. Spec. Top.-Phys. Educ. Res. **2**(1), 010101 (2006).
Ambrose, S. A. *et al.*, *How Learning Works: Seven Research-Based Principles for Smart Teaching* (John Wiley & Sons, 2010).
Anderson, C. W., *The Handbook of Research on Science Education*, edited by S. Abell and N. G. Lederman (Routledge, 2007), p. 3.
Bloom, B. S., *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*, 2nd ed. (Longman, 1956).
Brookes, D. T. and Etkina, E., Phys. Rev. Spec. Top. Phys. Educ. Res. **5**(1), 010110 (2009).
Brookes, D. T. and Etkina, E., Int. J. Sci. Educ. **37**(5–6), 759–779 (2015).
Brookes, D. T. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(1), 010103 (2021).
Chasteen, S. V. and Chattergoon, R., Phys. Rev. Phys. Educ. Res. **16**(2), 020164 (2020).
Chi, M., J. Learn. Sci. **6**, 271 (1997).
Cook, T. D. and Reichardt, C. S., *Qualitative and Quantitative Methods in Evaluation Research* (Sage, 1979).
Corpuz, E. D. and Rebello, N. S., Phys. Rev. Spec. Top. Phys. Educ. Res. **7**(2), 020103 (2011).
Creswell, J. W., *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, 2nd ed. (Pearson, 2005).
Daane, A. R. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **11**(1), 010109 (2015).
Dancy, M. H. and Beichner, R., Phys. Rev. Spec. Top. Phys. Educ. Res. **2**(1), 010104 (2006).
de Groot, A. D., *Thought and Choice in Chess* (Mouton, 1965).
Denzin, N. K. and Lincoln, Y. S., *Handbook of Qualitative Research*, 2nd ed. (Sage, 2000).
diSessa, A. A., *Constructivism in the Computer age*, edited by G. Forman and P. B. Pufall (Lawrence Erlbaum Associates, Inc., 1988), pp. 49–70.
DiSessa, A. A., Cogn. Instruct. **10**(2–3), 105–225 (1993).
DiSessa, A. A. and Sherin, B. L., Int. J. Sci. Educ. **20**(10), 1155–1191 (1998).
Erickson, F., *Handbook of Complementary Methods in Education Research*, 3rd ed., edited by J. L. Green *et al.* (Routledge, 2006), pp. 177–191.
Fosnot, C. T. and Perry, R. S., *Constructivism: Theory, Perspectives, and Practice*, 2nd ed., edited by C. Fosnot (Teachers College Press, 2005), pp. 8–38.
Franklin, S. V. and Hermsen, L. M., Phys. Rev. Spec. Top. Phys. Educ. Res. **10**(2), 020121 (2014).
Gee, J. P., *An Introduction to Discourse Analysis: Theory and Method* (Routledge, 2004).
Geertz, C., *The Interpretation of Cultures* (Basic Books, 1973), Chap. 1.
Glaser, B. G. and Strauss, A. L., *The Discovery of Grounded Theory: Strategies of Qualitative Research* (Aldine, Chicago, 1967).
Goldberg, F. M. and McDermott, L. C., Am. J. Phys. **55**(2), 108–119 (1987).
Green, J. L. *et al.*, *Handbook of Complementary Methods in Education Research*, 3rd ed. (Routledge, 2006).
Greeno, J. G. *et al.*, *Handbook of Educational Psychology*, edited by D. C. Berliner and R. C. Calfee (Macmillan, 1996), Vol. 77, pp. 15–46.
Gupta, A. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **10**(1), 010113 (2014).
Halloun, I. A. and Hestenes, D., Am. J. Phys. **53**(11), 1043–1055 (1985a).
Halloun, I. A. and Hestenes, D., Am. J. Phys. **53**(11), 1056–1065 (1985b).
Hamerski, P. C. *et al.*, Phys. Rev. Phys. Educ. Res. **18**, 020109 (2022).
Hammer, D. and Berland, L. K., J. Learn. Sci. **23**(1), 37–46 (2014).
Henderson, C., Am. J. Phys. **73**(8), 778–786 (2005).
Henderson, C. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **3**(2), 020110 (2007).
Hernandez, D. *et al.*, Int. J. STEM Educ. **8**(1), 1–17 (2021).
Heron, P. R., *Research on Physics Education* (pp. 341–350). IOS Press; *Research on Physics Education* (pp. 351–365). IOS Press (2004).

Howe, K. and Eisenhart, M., Educ. Res. **19**(4), 2–9 (1990).
Irving, P. W. and Sayre, E. C., Phys. Rev. Spec. Top. Phys. Educ. Res. **11**(2), 020120 (2015).
Jacobs, J. K. *et al.*, Field Methods **19**(3), 284–299 (2007).
Jordan, B. and Henderson, A., J. Learn. Sci. **4**(1), 39–103 (1995).
LeCompte, M. D. *et al.*, *Ethnography and Qualitative Design in Educational Research*, 2nd ed. (Academic, San Diego, 1993).
Lemke, J. L., *Talking Science: Language, Learning, and Values* (Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648, 1990), (hardback: ISBN-0-89391-565-3; paperback: ISBN-0-89391-566-1).
Lerner, G. H., *Conversation Analysis: Studies From the First Generation* (John Benjamins, 2004).
Lewis, C. and Rieman, J., Task-centered user interface design. *A practical introduction* (1993).
Meltzer, D. E., AIP Conf. Proc. **790**(1), 7–10 (2005).
Mestre, J. P. *et al.*, *Research on Physics Education, Proceedings of the International School of Physics "Enrico Fermi," Course CLVI* (IOS, 2004), p. 367.
Newell, A. and Simon, H. A., *Human Problem Solving* (Prentice-Hall, 1972).
Otero, V. K., *Proceedings of International School Physics Enrico Fermi, Course CLVI: Research on Physics Education*, edited by E. Redish and M. Vicentini (IOS Press, 2004), pp. 409–445.
Otero, V. K. *et al.*, *Reviews in PER* (American Association of Physics Teachers, College Park, MD, 2009), Vol. 2; available at http://www.per-central.org/items/detail.cfm?ID=9122.
Patton, M. Q., *Qualitative Evaluation and Research Methods* (Sage Publications, Inc, 1990).
Peshkin, A., Anthropol. Educ. Quarterly **16**(3), 214–224 (1985).
PhysPort, see https://www.physport.org/assessments/ (2022).
Posner, G. J. and Strike, K. A., *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice*, edited by R. A. Duschl and R. J. Hamilton (Suny Press, 1992), p. 147.
Posner, G. J. *et al.*, Sci. Educ. **66**(2), 211–227 (1982).
Psathas, G., *Phenomenological Sociology: Issues and Applications* (Wiley, NY, 1973).
Pulgar, J. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(1), 010120 (2021).
Rebello, N. S. *et al.*, *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. P. Mestre (Information Age Publishing, 2005), pp. 217–250.
Redish, E. F. *et al.*, Am. J. Phys. **66**(3), 212–224 (1998).
Reif, F. *et al.*, Am. J. Phys. **44**(3), 212–217 (1976).
Riessman, C. K., *Narrative Analysis* (Sage, 1993), Vol. 30.
Rogers, E., *Diffusion of Innovations* (Free Press, 2003).
Rumelhart, D. E., *Theoretical Issues in Reading Comprehension* (Routledge, 2017), pp. 33–58.
Ryan, R. M. and Deci, E. L., Am. Psychol. **55**(1), 68–78 (2000).
Sawada, D. *et al.*, School Sci. Math. **102**(6), 245–253 (2002).
Scherr, R. E., Phys. Rev. Spec. Top.-Phys. Educ. Res. **4**(1), 010101 (2008).
Schoenfeld, A. H., J. Learn. Sci. **2**(2), 179–214 (1992).
Scott, P. H. *et al.*, *The Handbook of Research on Science Education*, edited by S. Abell and N. Lederman (Routledge, 2007), p. 31.
Slotta, J. D. and Chi, M. T., Cogn. Instruct. **24**(2), 261–289 (2006).
Smith, D. C. and Neale, D., Teach. Teacher Educ. **5**(1), 1–20 (1989).
Spradley, J. P., *Participant Observation* (Holt, Rinehart, and Winson, 1980).
Stewart, D. W. *et al.*, *Focus Groups: Theory and Practice* (Sage Publications, Newbury Park, 1990).
Strauss, A. and Corbin, J., *Basics of Qualitative Research* (Sage Publications, 1990).
Strubbe, L. E. *et al.*, Phys. Rev. Phys. Educ. Res. **16**, 020105 (2020).
Thomas, D. R., Am. J. Eval. **27**(2), 237–246 (2006), p. 238.
Thornton, R. K., AIP Conf. Proc. **399**(1), 241–266 (1997).
Thornton, R. K., *Research on Physics Education* (IOS Press, 2004), pp. 591–601.
Thornton, R. K. and Sokoloff, D. R., Am. J. Phys. **66**(4), 338–352 (1998).
Trowbridge, D. E. and McDermott, L. C., Am. J. Phys. **48**(12), 1020–1028 (1980).
Tu, T. *et al.*, Phys. Rev. Phys. Educ. Res. **17**(2), 020142 (2021).
Tuminaro, J. and Redish, E. F., Phys. Rev. Spec. Top. Phys. Educ. Res. **3**(2), 020101 (2007).
van Zee, E. H. *et al.*, Sci. Educ. **89**(6), 1007–1042 (2005).
Vygotsky, L. S., *Thought and Language* (MIT Press, 1962).
Walsh, L. N. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **3**(2), 020108 (2007).
Werth, A. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 010129 (2022).
Wu, D. G. *et al.*, Phys. Rev. Phys. Educ. Res. **18**(1), 010121 (2022).
Yerushalmi, E. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **6**(2), 020108 (2010).
Zwickl, B. M. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **10**(1), 010120 (2014).

CHAPTER

26

RESEARCH-BASED TEACHING-LEARNING SEQUENCES IN PHYSICS EDUCATION: A RISING LINE OF RESEARCH

Jenaro Guisasola, Kristina Zuza, Paulo Sarriugarte, and Jaume Ametller

Guisasola, J., Zuza, K., Sarriugarte, P., and Ametller, J., "Research-based teaching-learning sequences in physics education: A rising line of research," in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 26-1–26-26.

26.1 INTRODUCTION

Research into physics teaching focuses on different scientific education goals such as understanding and improving content, promoting interest in science and its learning, and demonstrating the role of relationships between science, technology and society. At the center of all educational research lies the aim to improve student learning, as Hurd (1991) says "there is little reason to conduct research unless there is a pay-off in the classroom." Over the last few decades, research into physics teaching has helped improve the design and evaluation of instruction materials for physics courses. Many proposals have been based on a growing body of evidence which demonstrates that learning of scientific concepts can be improved as a result of research-based teaching (McDermott *et al.*, 1996; Fraser *et al.*, 2014; and Wieman, 2014).

The importance of basing education on evidence has become part of the educational administration discourse in many countries. Educational research is called upon to provide evidence of student learning in science education as well as in other educational areas, but in most cases, existing results are not considered when making choices and decisions on teaching the science curriculum. It is argued that research should offer a better base for choice and action than tradition or "professional knowledge,"

https://doi.org/10.1063/9780735425514_026

but no clear criteria have been determined for systematic improvement based on research. Over the last decade, the secondary and university physics teaching community has challenged ineffective practices (National Research Council, 2012; and Hazelkorn *et al.*, 2015) and provided evidence of innovations that offer opportunities for more effective teaching (Becerra-Labra *et al.*, 2012; and National Research Council, 2015). These demands are reflected in the teaching staff's teaching work. Most teachers who teach a discipline or group of disciplines devote much of their work to making decisions on how to plan and implement classes that develop the ideas, skills and attitudes of their students within a topic area. For science teaching research to directly influence educational practice, it must analyze and consider these types of decisions. It is not that we think that research on other scientific education objectives is less important. However, the line of research based on the systematic study of design, implementation and assessment of educational programs and materials can lead to support and stimulus for practices that have been demonstrated to be solidly conceived and effective when implemented (Cobb *et al.*, 2003).

Design-based research is a systematic study of the design, development and assessment of programs, processes and educational materials. This type of research might be valuable in improving educational interventions and understanding of the teaching process as well as student learning of the curriculum. In general, the research is carried out in real school environments, with all the complexity that accompanies teaching-learning in the classroom, raising significant methodological challenges for researchers in their efforts to maintain scientific rigor in their investigations. In this chapter, we consider some approaches taken to design and evaluate these educational materials and discuss existing problems and future perspectives.

Since the 1980s, various novel lessons, activities, and instructional strategies for teaching specific physics contents have been published that aim to connect research outcomes with the design of teaching materials. The literature refers to these teaching approaches as Teaching Learning sequences (henceforth TLS) (Meheut and Psillos, 2004). The TLS design research tradition has managed to improve existing teaching material by designing teaching activities based on research outcomes. This line of research focuses on the design and assessment of curricular products that cover the teaching and learning of a scientific topic (Heron *et al.*, 2005; and Savall *et al.*, 2019). These works include sequences of teaching activities to improve student learning on specific topics on a small scale (such as a few teaching sessions) or on a medium scale (a complete sequence of lessons on a specific topic), but they do not tackle a complete program on a large scale (for one or several academic years). One characteristic of TLS is its dual nature, both as an interventionist research activity in the classroom, and a set of teaching activities supported by research and empirically adapted to students' reasoning (Psillos and Kariotoglou, 2015). TLS design reflects how developing tools and learning environments relate to developing educational theory. This interrelation is a complex cyclical process where general education principles are applied to teaching specific topics in school contexts (Lijnse and Klaassen, 2004; and Juuti and Lavonen, 2006). Researchers have drawn up instruction frameworks to be used by the designers and teachers as interfaces between major theories and the needs associated with developing a TLS on specific topics.

There is a substantial body of research focused on the improvement of teaching materials which has based the design of TLS on existing science/physics education research results. This line of research aims at obtaining empirical results to, on the one hand, study the efficacy of particular TLS designs and, on the other hand, develop humble theories on classroom science teaching. These theories are "humble" in the sense that they are conclusions from analysing specific teaching-learning processes in a domain or field of the curriculum (Cobb *et al.*, 2003). These humble theories contribute to building a design theory which aims to explain why particular TLS design work provides insights on efficient TLS designs and suggests how they can be adapted to new circumstances. Therefore, like other methodologies, design-based research (henceforth DBR) constitutes a melting pot to generate and test general theories (Barab and Squire, 2004).

This chapter tackles the need to have a more systematic understanding of the achievements and challenges of TLS design research, including the problems related to effectively communicating these proposals to other researchers and to other teachers—to be used in different contexts. More specifically, we have two particular aims (see Fig. 26.1):

- To provide an in-depth account of the use of DBR as a methodology for the research-informed design, implementation and assessment of TLS sequences for secondary courses and introductory university-level courses.
- To suggest ways in which DBR can contribute to communicating with teachers the designed TLS so that they can use them effectively in their courses.

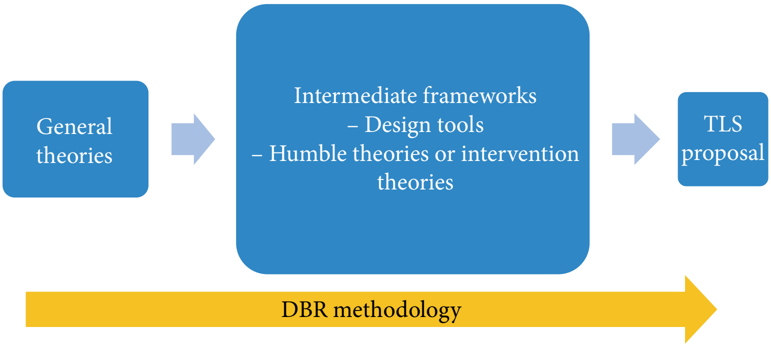


FIG. 26.1
Design based research methodology to build an intermediate framework to make explicit the design decisions of TLS.

In the following section, we present a brief review of the research on TLS design and implementation, the achievements obtained and the problems that remain. Section 26.3 describes Design Based Research as a research methodology. Below, we will discuss the concept of the design tool as an instrument that bridges the general teaching-learning theories and epistemology of science, and the specific instruction proposals contained in the TLS. The following section addresses the challenges of communicating TLS designs and their instructional frameworks to other teachers outside the design group and, specifically, the possible generalization of outcomes. We will conclude by discussing contributions from design-based research and new perspectives.

26.2 RESEARCH ON THE DESIGN AND EVALUATION OF PHYSICS TEACHING LEARNING SEQUENCES

From the early 1990s onwards, research into science teaching has helped develop teaching materials to improve teaching and learning in physics courses (Leach *et al.*, 2010; and Docktor and Mestre, 2014). Many of these teaching materials have been designed, implemented, assessed and re-designed with improvements to learning specific physics topics at a national and international level (Psillos and Kariotoglou, 2015).

Research into Teaching Learning Sequences (TLS) is rooted in two general theories of cognitive psychology. The first is based on Piaget's genetic epistemology (Piaget, 1971). This theory considers that students' understanding of scientific concepts and models can only be achieved by changing the concepts and ideas about natural phenomena students have before instruction; therefore, helping students make this conceptual change is the aim of science teaching. This implies that science teachers need to be aware of the students' previous ideas and to put forward classroom activities that promote conceptual change by guiding them to question and discard their previous ideas and to adopt the scientific ones (Posner *et al.*, 1982; and Osborne and Wittrock, 1983). Some of these theories have evolved to describe the students' "mental structures," while others look at mechanisms that encourage changes in the "mental structures" of individuals (Vosniadou, 1994; and Heron, 2003). Other theories focus on the knowledge-in-pieces cognitive framework, where the conceptual change is achieved by restructuring the students' primitive knowledge. According to this "knowledge in pieces" framework (DiSessa, 1993), students' ideas do not always correspond to stable conceptions but rather to spontaneous conceptions that respond to specific stimuli. This does not refer to "changing conceptions" but to revising previous ideas on scientific knowledge. Interest in science teaching research has brought in a second perspective based on Vygotsky's psychology. This perspective has been adopted by the teaching approaches based on the social constructivist learning theory (Tobin and Tippins, 1993). This theory indicates that cognition is not just in the person's mind but that it is a process that includes the student, the other students, the environment where they are learning and the learning activity.

Below, we present some TLS proposals for the last 4 decades. This is not an exhaustive review. We have selected proposals according to two criteria. On the one hand, these are proposals that attempt to make an explicit connection between existing research and theory to their proposals for teaching physics. On the other hand, these are proposals that have had a significant impact on terms of the number of teachers who have used their materials and/or the influence of other TLS researchers/designers.

In the late 1980s, the Children's Learning in Science-CLIS project was published and promoted by R. Driver and colleagues from the United Kingdom, based on Piaget's genetic epistemology (Driver *et al.*, 1985). The CLIS project design theory includes the following phases: (a) Stimulating students so that they present their own ideas and consider their classmates' ideas; (b) Introducing activities that interact with the students' previous ideas, restructuring their ideas towards the school view; and (c) Asking them to think about how their ideas have changed. This design theory had a great influence on interventionist research projects throughout Europe (Lijnse, 1994). Working from initial TLS design proposals, such as the CLIS project, the different TLS design approaches also included a description of student learning when working in the school context. This interest in the social contexts of teaching and learning led to research in Science Teaching as a social-constructivist perspective of learning that laid the foundation for the vast majority of TLS design proposals after CLIS (Tobin and Tippins, 1993).

In the 1990s, there was growing concern among research groups about applying research outcomes to classroom teaching. It was argued that a line of research in Science Teaching was needed that included "good teaching practice," not only considering the relevant educational research outcomes but also extending to science teaching research (Lijnse, 1995). Various science teaching research groups began and continue to develop "intermediate frameworks" that are explicitly based on one or several general theories to provide information on a particular practice, such as designing science lessons in school classrooms (Cobb *et al.*, 2003). These intermediate frameworks propose lines of action that are used to mediate between the contribution of general theories in epistemological and cognitive dimensions and the design process for TLS activities (Ruthven *et al.*, 2009).

The Physics Teaching Group at the University of Washington, led by Lillian McDermott (McDermott, 1991), has developed an instructional focus that is implemented in teaching sequences called "Tutorials" (McDermott *et al.*, 2003). Within this intermediate framework, physics teaching research, curriculum development and instruction are tightly bound in an iterative improvement cycle (Heron and McDermott, 1998; and Heron *et al.*, 2005). The instructional focus is "intentional" teaching: each activity in a TLS is carefully chosen with specific objectives and with views to the student's possible thought processes. The materials come with ideas to implement them in the classroom (McDermott *et al.*, 1996). Since publishing the first edition of "Tutorials" in 1998, tens of thousands of students have taken part in courses with tutorials at the University of Washington in dozens of colleges and universities in the United States and abroad, both in English and in translation.

The Science Teaching Group at the University of Valencia led by D. Gil and C. Furió proposes a didactic structure based on students constructing their knowledge as the result of the investigations to solve problems set by the teaching staff through a TLS that they call an "activity guide program" (Gil *et al.*, 1991; and Gil and Carrascosa, 1994). This approach looks at teaching as a guided program of tasks, where the students, directed by the teacher, address "problematic situations." In the problem-solving process, the students get involved in participation in scientific practices and at the same time improve their interest in learning. Resolving the task program guides the students through restructuring their prior knowledge in the pursuit of scientific knowledge (Gil *et al.*, 2002; and Guisasola *et al.*, 2008). This research has helped produce didactic materials, which have been implemented with comparatively good results (Savall *et al.*, 2019). An enormous number of students and teachers have followed their "Activity guide programs" in Spain and Ibero-America (Gil *et al.*, 2005)

P. Lijnse and K. Klaassen from the University of Utrecht, in their proposal that they call "Developmental Research" (Lijnse, 1995), describe a teaching strategy based on "setting problems" to develop TLSs on specific topics in the curriculum. This framework includes several phases: (a) Highlighting the interest and the reason for studying the topic in question; (b) Reducing the overall reason to specific needs for this particular topic; (c) Extending the students' knowledge in accordance with the overall reason and the specific knowledge needs; (d) Applying the knowledge to specific situations and problems; (e) Thinking about the need for orientation guided by the theory and developing more knowledge (Lijnse and Klaassen, 2004; and Kortland and Klaassen, 2010). This proposal has been widely followed by teachers in the Netherlands, in Primary and Secondary.

The group led by A. Tiberghien from the University of Lyon proposes an instructional focus that they call "Two Worlds" as an intermediate framework to inform the design of TLSs in Secondary physics (14–18 years old). It uses two fundamental orientations of the major theories of scientific epistemology and cognitive psychology. The former, based on the epistemology of science, treats modeling as a fundamental basis to build scientific knowledge. Secondly, based on Vygotsky's theory of learning, the science classroom is a place where students are invited to take part in an educational community where one of the teacher's roles is to replicate some of the knowledge and practices of the scientific community (Tiberghien, 1996). The design process follows the hypothesis that each time a person or a group explains or makes a prediction regarding the material world, in some way they are performing the modeling activity (Tiberghien, 2000; and Tiberghien *et al.*, 2009).

The group from the University of Leeds, led by J. Leach and P. Scott propose to design and assess TLS with a constructivist social focus (Ametller *et al.*, 2007). This focus uses design tools such as "Learning Demands" and the "Communicative Focus" from the teaching action (Leach *et al.*, 2006). The Learning Demand tool (Leach and Scott, 2002) attempts to identify the difference between the learning objectives required and the students' ideas, and the role that this difference might take when planning teaching strategies in the classroom. The "communicative focus" tool (Mortimer and Scott, 2003) focuses on discourse in the classroom and provides a perspective on how the instructor interacts with the students to develop ideas in the classroom's social aspect.

Duit and colleagues from the IPN–Leibniz Institute for Science and Mathematics Education (Kiel, Germany) proposed the "Educational Reconstruction" focus that includes a structure of recommendations to develop teaching-learning sequences. This approach combines the German tradition of science curriculum content analysis with the constructivist learning theory (Duit *et al.*, 2005, 2012). The model highlights the idea that scientific concepts and principles, conceptions of science and scientific research procedures, cannot be presented in a simplified way in teaching. Although, the structure of school scientific content is more basic from a scientific point of view, it is also richer in the sense that this content must be put into context for it to be understood by the students. This model considers that the tendency of some approaches to focus on changing the instruction methods, without also changing the traditional content structure, falls short (Fensham, 2001). The outcomes obtained in many research projects based on this didactic structure have demonstrated the usefulness of the recommendations and have become the main approach in TLS design in German-speaking countries.

Since the beginning of the century, Etkina and colleagues have proposed a framework that they call the Investigative Science Learning Environment-ISLE for educational intervention in physics classes (Brookes *et al.*, 2020; and Etkina *et al.*, 2021). The ISLE approach has two central goals: firstly, knowledge should not be separated from its discovery context; this leads to planning the students' work as an "epistemologically authentic" inquiry. Secondly, theoretical perspectives and design decisions must improve the students' interest in learning sciences and their perspective of being able to do so (Etkina *et al.*, 2008). Curricular materials based on ISLE come in different formats, including an algebra-based physics textbook, a guide for students and teachers, and materials on the Internet that are used at many U.S. universities (Etkina *et al.*, 2019a).

One would expect the discussion of how theoretical assumptions informed the design of TLS to be explicitly shown, but in the programs mentioned in the brief review provided in the previous paragraphs of this section, this is often not the case. This is partly due to the great breadth of the investigations that cover a wide variety of school interventions with different specific aspects (learning, cognitive development, teaching strategies, interactions in the classroom, etc.). However, the majority of the aforementioned proposals share some suppositions on teaching and learning sciences (see Table 26.1) that emerge from science education research that include (a) considering students' alternative ideas from a constructivist perspective of learning; (b) considering that learning is not a simple process of transferring knowledge from the teacher to the student, and therefore teaching needs to provide opportunities to clarify meanings and practice scientific skills; (c) designing activities depending on research outcomes from a social-constructivist perspective of the teaching-learning process in the classroom; and (d) presenting (usually) proof of student learning. These shared suppositions can be considered at *large grain sizes*. The term "grain size" is used to indicate the level of detail used to describe the teaching contents and the pedagogic approximations in relation to the specific content (Leach and Scott, 2002). By using a large grain size, we are referring to general ideas on the process of

Table 26.1

General theories and proposal product in the revised approaches to the design of teaching learning sequences in Physics.

Lines of work on TLS design (key reference about the framework for further details see the references of this section)	General theories	Main design elements
Children's Learning in Science-CLIS (Driver *et al.*, 1985)	Piagetian perspective on knowledge construction; "classic" conceptual change theory	The teaching task for eliciting students' ideas and restructuring them.
Teaching physics by inquiry (McDermott *et al.*, 2003)	Piagetian perspective on knowledge construction; current conceptual change theory; difficulty framework	Program of activities
Teaching and learning by guided research (Gil *et al.*, 1991)	Piagetian perspective on knowledge construction and Laudan's perspective of the epistemology of science	Activity guide program
Developmental Research (Lijnse, 1995)	Social constructivist perspective of learning; psychological theories about the attitude and interest of students	Problem posing
Two Worlds (Tiberghien, 1996)	Vygotskian perspective on learning; Bachelard's epistemology of science; Vygotskian perspective on learning	Design Modeling relations and Knowledge distance
Design Brief (Leach *et al.*, 2010).	Sociocultural account of meaning-making on the social plane; Realist ontology	Learning demand; and Communicative approach
Educational reconstruction (Duit *et al.*, 2005)	Social constructivist perspective of learning; Epistemology of science	Epistemological analysis of contents at the educational level
Investigative Science Learning Environment-ISLE (Etkina *et al.*, 2021)	Socio cultural perspective of learning; Teaching as intentional activity; Nature of Science	Teaching by "epistemologically authentic" inquiry

teaching and learning sciences. However, theoretical elements with a large grain size are not enough. More specific information—*fine grain size*—is required on the specific scientific content and how to address its teaching and learning. For instance, although it is useful to know that students' previous ideas on natural phenomena influence their learning of scientific concepts and models, it is more useful to have information on the specific problems that students have when learning a particular science topic (Etkina *et al.*, 2019b). Cobb and Gravemeijer (2008) mention that it seems quite difficult to articulate specific instruction theories in the absence of sequences of activities and associated resources to support the learning. As we have seen, progress has been made in articulating intermediate frameworks that relate major theories and design decisions, but there are still significant gaps in the area of design-based research. Specifically:

- Explicit articulation of the TLS design methodology is required that includes the theoretical commitment regarding the research and how this leads to design tools that specify the TLS implementation and assessment.
- Although many approaches talk about "active teaching" or "active learning," there is frequently no detailed specification of the teaching strategies.
- In relation to the evaluation of the proposed approach, a broad evaluation procedure is not usually presented that might go beyond the learning achieved by the students. This implies that attention is often not paid to the sequence re-design and refining process. The lack of these explicit descriptions inhibits cumulative progress in the teaching.

After analysing the achievements and persistent problems in the research on teaching proposals on specific topics in the curriculum and the TLS design, in the following section, we describe attempts to overcome these problems through the movement for "design-based research" (DBR).

26.3 DESIGN-BASED RESEARCH: A RESEARCH METHODOLOGY ON TEACHING AND LEARNING SPECIFIC SCIENTIFIC CONTENT

In the introduction to the special issue of the *Educational Psychologist journal*, Sandoval and Bell (2004) mention the tension in educational research between producing universally applicable and reproducible knowledge and specific studies that address teaching certain topics on the curriculum with complex educational interventions in specific environments. Over the last few decades, work programs have been developed known as Design-Based Research (DBR), where theoretical and empirical knowledge on learning is systematically used with a fine grain size in the design, implementation and evaluation process for teaching materials (Design-Based Research Collective, 2003).

Design-Based Research stands out from other research into Science Teaching due to two main features. Firstly, DBR focuses on understanding disorder in the practice of classroom teaching, taking the context as a central part of the research and not as a variable that can be trivialized. The DBR approach involves the characterization of situations (as opposed to the control variables) and developing a humble intervention theory for the classroom that characterizes the design in practice (instead of simply proving a hypothesis) (van den Akker, 1999; Barab and Squire, 2004; and Bell, 2004). The second main characteristic is that DBR implies the production of demonstrable changes at a local level. That is, design research takes into account the characteristics of local contexts to design TLS which are feasible for that context.

There is consensus in the literature in terms of defining DBR as a research methodology that integrates design and general theories to generate useful products and humble theories to solve teaching problems with specific topics (van den Akker, 1999; Design-based Research Collective, 2003; Juuti and Lavonen, 2006; and Alghamdi and Li, 2013). The methodology does not assume a specific educational theory that sustains it or specific tools for any phase, which gives education researchers considerable freedom on how to implement DBR (Reeves, 2006; and Easterday *et al.*, 2014). Most authors agree that a DBR project should be developed through design, implementation, assessment and redesign cycles (McKenney and Reeves, 2018). Each of these parts will be described below, considering that these phases are not a linear sequence but are configured in an interactive process.

26.3.1 Design process

The design process consists of different elements that include the educational context, the specific content of the topic to be taught, the perspectives from the literature on the teaching and learning difficulties, lessons from the appropriate teaching strategies to meet the defined learning objectives, and design of the activities from the sequence.

26.3.1.1 Educational context

Because the design-based research takes place in real teaching contexts, most contextual school elements should be explicitly identified, such as the type of target students for this TLS and the teachers' professional knowledge and experience, that will limit the scope of the TLS. This phase is key in the design, as the literature shows (see Sec. 26.2) that one of the difficulties of the TLS design process is the lack of clarity in the contextual factors involved in the sequence.

26.3.1.2 Definition of the teaching-learning goals

One claim in the literature for an effective design process (see Sec. 26.2) is the need to define the specific contents of the topic and justify this definition. This implies an epistemological analysis of the contents in the school curriculum context. This involves providing epistemological evidence from the discipline that allows us to justify the choice of the teaching and learning goals, avoiding the definitions based on the idiosyncrasy of the designers or on the non-explicitly traditional curricular elections. The objective-defining process also includes reviewing existing information on known difficulties in learning the topic and existing teaching solutions to overcome them.

In the light of the results of the review of previous studies and epistemological analysis, the designers determine their learning goals and the indicators that will be used to evaluate their achievement. It is crucial to define the learning goals clearly and explicitly if we want the results of the TLS assessment to be useful in future designs, although some freedom is permitted when defining these goals, in addition to the contextual factors that might limit them. Naturally, this would mean that different TLS designed around the same focus would interact constructively to improve the definition of the design decisions. This phase is hard to express (McKenney and Reeves, 2018), and as we will see in Sec. 26.4, it requires elements such as the Design Tools that systematize the justified definition of the TLS contents and the learning demands required from the students.

26.3.1.3 Design solution

In this phase, the designers outline a possible solution to take the defined objectives to the classroom. This is a phase to produce teaching materials and define the assessment guidelines in accordance with the defined learning objectives. This phase is where the TLS product is presented as the solution for the relationship determined between epistemology, learning, and teaching, adapted to the analysis of teaching sequences on a specific topic. One important characteristic of this solution phase is that many key decisions have a very fine grain size, related to specific aspects of the content that is going to be taught or the teaching focus.

The design with the activities would include drawing up guides for teachers who provide information on the design decisions related to the information sources selected by the designers and guidelines on the teaching practice when using the activities included in the TLS (Ametller *et al.*, 2007). In both cases, the aim is to inform teachers about the planned TLS implementation so that it is consistent with their key points (Pintó, 2005).

26.3.2 Implementation: Teaching experiments

Classroom implementation of the TLS aims to investigate the hypothesis that the design will lead to better student learning, judging by the chosen assessment strategies. This phase can be considered a "teaching experiment." As Cobb *et al.* (2003) state, "a primary goal for a design experiment is to improve the initial design by testing and revising conjectures as informed by the ongoing analysis of both the students' reasoning and the learning environment" (p. 11). In particular, the teacher or the teaching team might feel the need to adjust the TLS as the students go along. These changes might be due to incidents in the classroom such as students who take approaches that were not planned, activities that are too difficult, etc. These adjustments are generally not accepted in the comparative experimental research, although in DBR methodology, changes are made in the TLS to create optimum conditions and are considered part of the body of data.

26.3.3 Retrospective analysis: Evaluation and redesign

This stage evaluates the proposed design to assess its efficiency in relation to the objectives that were addressed. One central challenge of the retrospective analysis is that the conclusions should be reliable and so, the criteria should be explained along with the types of evidence used when making inferences. However, the DBR methodology does not determine the type of analysis tool that should be used. These tools should be chosen by the researchers in accordance with the aspects that are evaluated. However, the DBR highlights the need for multiple and convergent evaluation designs to be included that clearly set out the different aspects of the TLS evaluation (Nieveen, 2009). In most design projects, TLS efficacy is tested in two dimensions:

a. Sequence quality analysis that includes (a1) problems related to the clarity of the activities that students must perform; (a2) problems related to the time required to complete the sequence; and (a.3) problems derived from a proposal that differs from the traditional content sequence.
b. Learning outcomes analysis that includes student comprehension: (b1) of the concepts, models and theories; and (b2) development of scientific skills.

To evaluate the different elements in the two dimensions, both qualitative and quantitative research methodologies are used. Because the design research is exploratory, the aim of the evaluation is not only to obtain quantitative data on the learning achieved but also to determine whether the retrospective analysis is useful to overcome some of the difficulties identified in the TLS design and redesign. For example, qualitative methodologies are used such as the “teacher’s diary” (Carr and Kemmis, 1986), the “student workbook” (Leslie-Pelecky, 2000), or the “External assessor’s report.” The learning analysis dimension mainly uses tools from the quantitative investigation, such as questionnaires with open or multiple questions, to evaluate the comprehension of the concepts and the students’ theories. In addition, they are usually complemented by student interview analysis.

The evaluation results can influence aspects of the TLS redesign, such as re-writing parts of the text, modifying the analogies used, analyzing whether the representations have been understood and the general focus of the TLS. This improves the probability of finding an effective design that can be subsequently verified through the final evaluation (Etkina *et al.*, 2009; Guisasola *et al.*, 2017; and Zuza *et al.*, 2020).

Although the redesign of a TLS by analyzing its implementation is fundamental in DBR, the efficacy with which this re-design can be achieved is influenced by the quality of the original design and the clarity and coherence of the decisions that were made. Ensuring original robust designs that work properly requires “design tools” to be developed that can identify and address specific aspects of the TLS and help both the initial formulation of the design and its subsequent refinement (Kelly *et al.*, 2008). Consequently, design tools have been developed to influence the fine details of the design solution that we will mention in the next section.

In the DBR methodology, the evaluation of both the teaching-learning sequence itself and the learning achieved by the students in relation to the proposed learning objectives and the school context is a fundamental part of the TLS approach. Without showing results of the implementation of a TLS proposal, it is unlikely that teachers and researchers will accept and embrace the approach for implementation. However, the aim of this article is to show that the DBR methodology is useful for making the development of research-based TLSs more transparent. In addition to describing the elements that make up DBR, the chapter also describes several “design tools” that could provide support for teachers of physics and physics education researchers in developing TLSs. Therefore, in this manuscript, we do not develop the evaluation section. Our way of adapting the DBR methodology to the evaluation section can be seen in the previously published research (Guisasola *et al.*, 2017, 2021; and Zuza *et al.*, 2020).

26.4 DEFINING DESIGN TOOLS FOR ARTICULATING INTERMEDIATE FRAMEWORKS

The idea of a "design tool" as a concept that is used to design teaching based on theoretical knowledge and empirical results is rooted in the design research line that takes engineering as a metaphor for scientific education (Hjalmarson and Lesh, 2008). The expression "didactic engineering," as explained in Artigue (1992), emerged to name a form of didactic work that is comparable with the work of an engineer. While engineers base their work on knowledge of scientific theory, they are obliged to work with more complex purposes, and therefore, to manage problems that science cannot or will not address yet. This way of working in engineering stands as an analogy for labeling the work of research projects on educational interventions in class. In fact, the expression "didactic engineering" has become polysemic to designate both productions for the teaching derived from or based on the research, and a specific research methodology based on experimentation in the classroom (Artigue, 2014).

This chapter uses the term "design tools" to refer to constructs that provide information on TLS design decisions based on educational theories and empirical results of science education research (Ametller *et al.*, 2007; and Ruthven *et al.*, 2009). Design tools are related to humble theories on classroom intervention, but while humble theories contextually enrich the intermediate frameworks, design tools help to systematize the design process and are connected with the design activity (Cobb *et al.*, 2003). Consequently, this section seeks to help present and/or develop several theoretically informed design tools for TLS design. We do this based on international research traditions in science teaching over the last 40 years described in Sec. 26.2.

After we discuss the design tool construct, we consider a number of specific design tools. These tools will help designers mobilize theoretical and empirical research insights into TLS design elements. These elements can cover a wide range of aspects, from orientations on defining learning goals or organizing group work in classrooms to the proposal of topic-specific activities. In this chapter, we will address the design tools focused on TLS elements that can be based on the research. Nothing like an algorithm should be expected from the design tools, as a certain degree of professional knowledge should be applied in the process. The virtue of the design tools lies in clearly setting out this process and framing the field of application of the professional knowledge and criteria to make it easier for designers to document, evaluate, and discuss the emergence of the design's constituent elements.

Therefore, we should start by referring to the elements or aspects for which there is sufficient information and that we can argue are sufficiently important to justify informed research decisions.

We will pinpoint approaches to analyze the cognitive and epistemological dimensions of a specific domain to report both on its design for the sequence and how it works in implementation. Three design tools are presented below that are used in TLS design with wide consensus in the PER community.

26.4.1 “Epistemological-ontological content analysis” tool

One central element in the choice and justification of the concepts, models and theories to be taught on a specific topic of physics is what we call the epistemological and ontological analysis tool at the chosen educational level. This tool was developed to help identify the learning goals for science teaching at fine grain size. This refers to connecting the knowledge from the theoretical framework of physics that includes not only theoretical concepts and models but also ontological and epistemological elements from the Nature of Science (NOS), with its teaching design in the TLS.

The *epistemological-ontological content analysis* tool considers knowledge of physics as the outcome of an arduous process of problem solving and rigorous testing of the initial hypotheses (Nerssesian, 1995). A detailed analysis of the historical events related to science shows that sciences are disciplines where change is more the rule than the exception (Kuhn, 1984). The current consensus is that scientific education requires not only knowledge of concepts and theories but also development of scientific skills and knowledge of the nature of science. The structure of science, the nature of scientific practice and the validation of the scientists’ judgments are some of the areas where the characteristics of the Nature of Science can enrich the teaching and learning aims for Science (Wandersee, 1992; Duschl, 2000; McComas *et al.,* 2000; Rudge and Howe, 2004; Hodson, 2014; and Matthews, 2014). The consideration by researchers and teachers of the “discontinuities” (that is, sharp changes) between different physical models through history can help them clarify, explain and explore the physics concepts and better understand the key aspects of a topic. For example, a study of the developments leading from the impetus model of force to Newton’s interactive idea of force can help students better understand the conceptual obstacles that must be overcome in order to understand Newtonian force theory. The analysis of this “discontinuity” in the development of the theory can lead to an understanding of the difficulties of the students.

Research on teaching-learning specific content shows that the lack of an appropriate epistemological analysis can be related to inaccurate or incorrect conceptual approaches that can be present even when the general goal is coherent with the study plan. See, among others, “the clarification and analysis of science subject matter” proposed by Duit *et al.* (2012) or “analysis of the content knowledge” proposed by Tiberghien *et al.* (2009). As a summary, the ontological-epistemological analysis of the curriculum content as a design tool provides a way of clearly identifying the key scientific ideas that guide the conceptual content to be taught and from there define the support objectives at a specific level to begin to describe the fundamental aspects of the TLS program (Hodson, 2014; and Guisasola *et al.,* 2017).

To exemplify this, it will show how the curriculum *Epistemological-ontological content analysis tool* is used to justify the content design for a teaching-learning sequence for Newton’s Laws for Secondary

Table 26.2

Relationship between the key epistemological elements and the learning objectives. The curriculum analysis is based on a Teaching by Guided Problem Solving (GPS) approach that guide students on using scientific practice such as knowing how to approach problems, make hypotheses, compile data, and make arguments based on empirical evidence.

Epistemological key components of the Newton's laws	Physics knowledge to be taught in the TLS
The concept of force: K1. Force measures the interaction between bodies A and B. Self-force does not exist. K2. The force exerted by body A has exactly the same magnitude as the force exerted by body B, and they are simultaneous (Newton's 3rd Law). Both forces have the same line of action and yet have opposite directions. Vector nature of the force.	*Learning objectives:* For a given system: O.1. Recognize exerted forces and each of the "agents," or interacting bodies responsible for the mutual forces, and draw a force diagram.
Action of the force K3. Newton's 2nd Law relates the force exerted on a body and the change in its state of movement (the acceleration; vector magnitude) that it acquires through the inertial mass magnitude. K4. As a consequence of the above, the inertial mass is a property of the body that measures its resistance to the change in its state of movement (Newton's 1st Law).	*Learning objectives* 0.2. Applies Newton's second law: O.2.1. Writes the equation O.2.2. Calculates the resultant force and acceleration
Force and change of movement: K5. The acceleration resulting from the action of forces can be decomposed into two orthogonal components: tangential acceleration (changes the speed) and centripetal acceleration (changes the direction of velocity).	*Learning objective* O.3. Recognize that the resulting acceleration can change both the direction of velocity and the speed. Calculate the normal (centripetal) component of the acceleration. O.4. Shows problem solving skills

Teaching (16–18 years old) (Guisasola *et al.*, 2019). The example shown involves determining the relationship between the key elements of the epistemological and ontological analysis of the theoretical physics framework and the learning objectives for the sequence (see Table 26.2).

The analysis of the context of discovering Newton's laws and problems that scientists had to overcome as far as the current definition of the Force concept in classical mechanics, for an education level of 16–18 years old, tells us that Force is a quantity, having both magnitude and direction, that measures the interaction between two bodies. There is no force without the presence of at least two bodies that interact; furthermore, this interaction must be mutual, occur in each of the bodies at the same time have the same magnitude and act along the same line, although in the opposite direction. This theoretical construct is the start of the nature of force in classical physics (what is known as Newton's Third Law). If there are more than two bodies, the Superposition Principle is met with two-by-two interactions. This principle shows that the "directional" nature of the interactions and their quantitative application requires an understanding of the vectorial nature of the Force concept. If the resulting forces exerted on a system are not balanced, we can suppose that there will be a change in movement. In the development of Newton's Laws, this is how the relationship was defined between force and acceleration (Newton's 2nd Law) (Coelho 2010). In classical physics, inertial mass is considered as the property of the bodies when they oppose the change of movement or their capacity to attract each other (Newton's 1st law) (Ellis, 1962). The brief statement of Newton's first law seems simple; however, its meaning represents conceptual clarification discussions and overcoming different comprehension difficulties (Marquit, 1990). Several authors show that the first law can be represented in a temporary and quantitative form. The temporary form refers to remaining in the state of movement if no external force acts and the quantitative form tackles the change in the state of movement of a body when an external force is applied. Both representations are complementary, and the first law must not be simplified in just one of the representations. The quantitative form particularly reflects a conceptual change in the comprehension of the movement, eliminating the rest-movement opposition (Galili and Tseitlin, 2003).

The program proposal that emerges from the epistemological analysis indicates that the nature of the Newtonian force includes the pair interaction, i.e., the third law, as a fundamental element of force. This clash with a teaching approach centered on the ascending order of numbering of Newton's Laws. In particular, using confusing language for the third law with expressions such as "For every action there is an equal and opposite reaction" that seems to be a synonym for "cause and effect" as if the "reaction" is preceded by an "action." All this leads us to suspect that the explanatory development of Newton's laws may differ from the numerical sequence given by Newton to his laws (see Table 26.2).

The example clearly shows that this tool is intended to move from the epistemological components to the conceptual learning objectives. The onto-epistemological analysis helps us clarify the conceptual components that are required to build knowledge on a topic and the logical and conceptual relationships between them. This information enables decisions on the key concepts and the order in which they should be worked on. This information can be used to specify the curricular goals in a way that is grounded in learning objectives. This still lacks the didactic transposition and the activities to work on the outcome of the transposition.

The learning objectives of Table 26.2 take into account the contextual analysis of the proposal, which is addressed to a first course of Newtonian physics in High School. In this sense, in the proposed TLS, we always choose inertial reference systems and leave the discussion of the influence of the reference system on the validity of Newton's laws for a second part of the topic of Newton's laws.

26.4.2 Learning demands tool

This tool explains the conceptual, ontological and epistemological differences between the knowledge to be taught and the students' knowledge, providing a detailed map of the distance between the two. The *Learning Demands* tool is based on social-constructivism (Leach and Scott, 2002), particularly the concepts of prior or alternative conceptions, the zone of proximal development (ZPD) and conceptual change. The tool compares two pieces of knowledge from the material world: knowledge from school science and students' knowledge.

The students' ideas can generally be found in the extensive research literature on prior conceptions or, when this is not available for a particular topic or context, the team of designers must perform an investigation to describe them. Once both inputs have been identified, the designers must determine the conceptual, ontological and epistemological differences between the starting point and the final objective of the planned learning process, defined for each school context in the learning objectives that are derived from specifying the curriculum through onto-epistemological analysis. On this point, the designers describe the type of gap (ontological or epistemological) and, from the existing experimental results and their professional knowledge, they decide on the "measurement" of this gap as an expression of how difficult it is for students to overcome it.

Now we will illustrate the identification of the learning demands for the same topic that we used in the epistemological and ontological analysis tool. There is a broad bibliography on students' difficulties in learning the concept of force and Newton's Laws. The bibliography indicates that the students show difficulties in understanding the concept of force as an interaction between two bodies and its vector aspect. Furthermore, students' explanations tend to be based on a cause-and-effect sequence with a body that exerts the force and another that receives it in a time sequence. The students tend to find different explanations to explain the cause of the movement of the bodies and the state of rest and frequently do not relate the state of rest to a specific case of the force and movement relations (Champagne *et al.*, 1980; Clement, 1982; Hestenes *et al.*, 1992; Hewitt, 2002; Duit, 2009; Barniol and Zavala, 2014; and Andhika *et al.*, 2016). A summary of the difficulties found and their relationship with the learning objectives can be seen in Table 26.3.

The *Learning Demands* tool provides us with a qualitative measurement of the gap (low, medium, high) that mainly depends on the empirical studies regarding student difficulties but also on the teaching team's experience and the school context where the sequence takes place. In this respect, the tool has a professional knowledge component frequently based on experimental results. This qualitative classification of the degree of difficulty, however, is also based on three general characteristics of students' conceptions of difficulty (Taber, 2017): (1) Degree of Inconsistency Scientific Models; (2) Degree of Connectedness: how relevant is the inconsistency in other areas of the curriculum; (3) Degree of Commonality: How common a particular inconsistency is among students. Just as the onto-epistemological analysis becomes a design tool because it allows us to pass from the curriculum to the learning goals, determining the gap helps us orient aspects of the teaching in the socio-constructivist framework, suggesting types of activities and strategies and timing for each objective.

26.4.3 Tools for "Staging a Teaching sequence"

The application of the didactic tools above provides fine grain information for TLS design. This information makes it possible to define a TLS framework that includes content, sequencing, goals, foreseeable difficulties and indications of the attention they will require. Working from this information, two phases remain to obtain the final TLS. Firstly, pedagogic strategies should be chosen to address the learning objectives in accordance with the specific information obtained by using the design tools

Table 26.3

Using the Learning Demand tool to analyze the gap between students' difficulties and learning objectives in the Newton's law sequence at Secondary Education level.

Learning objectives (Physics knowledge to be taught in the Sequence)	Knowledge gap (according to empirical studies and teaching experience)	Learning difficulties (Students' existing physics knowledge)
For a given system: O.1. Recognizes exerted forces (and each agent) and draws a force diagram.	High	D.1. Identifying as forces only those that produce "changes" D.2. Larger-mass objects exert larger forces D.3. Work with vector magnitudes D.4. Comprehension of graphs
0.2. Applies Newton's second law: O.2.1. Writes the equation O.2.2. Calculates the resultant force and acceleration	High Medium	D.5.The idea of Impetus, D.6. Larger-mass objects exert larger forces, D.7. Friction forces make bodies stop, D.8.The last force dominates the motion of the body, D.9.Force and velocity are proportional, D.3.Work with vector magnitudes, D.4. Comprehension of graphs
O3. Recognizes resulting acceleration can change both the direction of velocity and the speed. Calculate the normal (centripetal) component of the acceleration., O.4. Shows problem solving skills	Medium Medium	D.1. Identifying as forces only those that produce "changes," D.3.Work with vector magnitudes, D.4. Comprehension of graphs D.5. The idea of Impetus D.7. Friction forces make bodies stop, *General difficulties in using scientific skills*, such as qualitative understanding of the problems, making hypothesis, elaborate solving strategies, analysis of results.

above. The selection of the pedagogic tools must be justified in each case working from the theory (general and humble) and the relevant empirical outcomes. By including the pedagogic strategies, the TLS framework becomes a "design brief" (Ametller *et al.*, 2007). The last step in the design involves specifying the activities that are clearly set out by the pedagogic tools. This last step is where the teacher's knowledge and the adaptation to the specific teaching context carry the most weight.

The step from the "design brief" to the activity design phase can be made certain Staging tools such as tutorials or guided problem solving (GPS). These design tools make it possible to inform the process of passing from the framework grounded in TLS to a sequence of activities of the type that defines a certain educational proposal. As a set, these tools are a family of design tools that we call "Staging of teaching sequences."

We present an example of this type of tool within the pedagogic strategy that we call *Teaching by Guided Problem Solving (GPS)* (Barell, 2006; Guisasola *et al.*, 2008; and Zuza *et al.*, 2014) for the same topic as used for the previous tools. In accordance with the commitments to the social constructivist theory of learning and the epistemology of Science, we set up a teaching trajectory of the TLS based on a structure of driving-problems that guide students' learning (first level of specification). In order to solve each driving problem, a set of activities is used as scaffolding to help students in the resolution of each problem (second level of specification). We use a pedagogic strategy of teaching that we call Teaching by Guided Problem Solving (GPS) (Guisasola *et al.*, 2020). In this teaching strategy, the students are not explicitly confronted with the alternative ideas they may have; instead, they are given the opportunity to modify or refine their ideas in the light of guided questions for solving the posed driving problem. For the resolution of the driving problems, the class is divided into small working groups (3–4 students) that develop a preliminary inquiry. In each driving problem, there are a number of activities that guide students to solve the problem. Students discuss and work out activities in their small group. Then, group answers are pooled and analyzed by the students and the teacher, coming to a reasoned consensus. During group work, the teacher's role is to encourage and guide students, to question their answers and to make them think about it or provide additional information if necessary. The activities are accompanied by a teacher's guide which discusses the relationship of the activity to the learning objectives and justifies the type of teaching technique (rapid-response system, flipped classroom) most appropriate for each activity. The type of active teaching technique used aims to give students opportunities to use scientific practices and evidence-based reasoning to communicate their ideas (Jiménez-Alexandre *et al.*, 2000; and Verdu and Martinez-Torregrosa, 2004). The traditional authority and novice roles blur together as the students work in cooperative teams to solve problems that have already been solved (as opposed to original research), watched over by a teacher who is familiar with the solution. In our approach, the teacher plays an essential role in posing problems and in guiding both their resolution and the learning process.

We show some examples of how we use the tools for staging the TLS. At the first level of concreteness, with regard to the structured learning path based on driving problems, we pose two driving problems to the students:

1. What happens when forces are exerted? What is a force?
    1.1 How can we represent and measure forces?
2. What is the relationship between forces and motion?
    2.1 How can we measure the relationship?

Table 26.4 shows the outline of the construction of the TLS activities according to the defined objectives, learning demands and teaching strategy.

The activities mentioned in Table 26.4 are shown below as examples (Table 26.5).

In Sec. 24.1 of the TLS, the students are given activities that allow them to answer the driving problem 1. The example in Fig. 26.1 shows only the first three activities that have as a general objective that the students become familiar with phenomena involving forces and begin to understand the Newtonian meaning of force and its vector representation (O1). In activities A.1 and A.2, the concept of force as an interaction and Newton's third principle are explicitly worked on. As we commented previously, in the

Table 26.4

Construction of the TLS activities with the help of tools for staging the TLS.

Driving problems	Learning objectives (demands)	Strategies to foster learning GPS	TLS. Activities and comments
What happens when forces are exerted? What is a force? How can we represent and measure forces?	O.1. Recognizes exerted forces (and each agent) and draws a force diagram.	• Familiarize students with analysing phenomena of forces • Propose hypotheses on the role played by the force in different situations and test then by evidence • Teaching Technique: Rapid response system (Socratic): Discussing the hypothesis	Activities to build an explanatory model of the force as interaction by contact or at distance. A1, A2: Activities to understand the concept of force as interaction (action and reaction principle) A.1, A.2. A.3: Activities to apply the concept of force as a vector and force diagram

Table 26.5

An example of the first three activities presented to students as scaffolding to start them thinking about and solving the driving problem 1.

I. Force as interaction
A.1. A box on a table. Choose the option that best describes the forces acting in this situation:
a) Gravitational force exerted by the earth at the center of the box downwards.
b) No forces are acting
c) Gravitational force exerted by the earth on the box downwards and the force exerted by the table in the center of the box upwards.
Explain your choice and draw the force diagram.
A.2. A bird flying. Choose the correct option:
a) The bird does not put any force on the Earth.
b) The same force that the Earth exerts on the bird, the bird exerts on the Earth.
c) The bird exerts a force on the Earth, but not of the same magnitude as the Earth exerts on the bird.
Explain your choice and draw the force diagram.
A.3. A person throws a ball forward. What force acts on the ball while it is still in the air?
a) The force of gravity exerted by the Earth and the force exerted by the person.
b) The force of gravity exerted by the Earth and the friction exerted by the air.
c) The force of gravity exerted by the Earth, the force exerted by the person and the friction exerted by the air.
Explain your choice and draw the force diagram.

teaching by GPS, for the resolution of activities, the students work in groups of 3–4 students and develop a preliminary solution of the activity based on their previous knowledge and justifying their response.

In the case of A.1. a situation is presented, which may present difficulties of the D.1 and D.3 types (Table 26.3). In fact, option (a) and option (b) are "alternative" explanations which the literature shows are used by a significant number of students. They involve students discussing in groups and then expressing their opinion individually through the Socratic quick response program. In sharing the explanations of the groups with the class, the teacher insists on the balance of forces, which implies that the body is at rest on the table, on the nature of each force and on the vector representation. The process is repeated with activities A.2 and A.3. In A.2., two distractors are proposed in options (a) and (c) related to difficulties D.2., D.3. and D.4 (Table 26.3). In activity A.3. is a question widely used in the literature with the aim of discussing the instantaneous nature of Newtonian force.

The design of the specific tasks and the way that they are going to be presented to the students is guided by Staging a Teaching Sequence tool, although the design of activities is not unique and will depend on the school curriculum, the pedagogic content knowledge of the designers and the school context.

26.5 EXCHANGE OF SUCCESSFUL EDUCATIONAL PRACTICES AND PRODUCTS

One important question that DBR must answer is what is considered acceptable evidence. In educational research, a solid methodological argument must be related to valid, reliable and useful topics as well as the different contexts where researchers think that the results could be generalized. In DBR, to demonstrate that the validity and reliability criteria have been met, similar methods are used as in other research in education (Cohen *et al.*, 2013). While the generalization and external validity criteria are usually mentioned less frequently when evaluating the impact of the research results, aspects regarding the dissemination of best practices have become relevant over the last few decades in science teaching proposals, as so-called educational interventions must prove that they are effective in the classroom context (Testa *et al.*, 2020).

In relation to the generalization of the design research, it is one thing to demonstrate the learning gains or show that statistical differences have been achieved between control and experimental groups, and quite another thing to demonstrate the usefulness of the didactic structure that guides the proposal design. As we have mentioned, the design research is carried out in local cultural contexts and, depending on the topic and the educational level, it is difficult to replicate the findings of others. One determining factor when disseminating the TLS and its application is the teacher's understanding of the "main ideas" from the didactic material and how this is implemented in the classroom. By "main ideas," we mean substantial relationships between the learning objectives and the design of activities that offer students opportunities to learn these objectives using scientific practice. These relationships are rarely obvious in the usual instruction materials. These didactic materials with broad focus points such as "inquiry," "hands-on work," or "standards-based teaching" are usually poorly defined and do not help the teacher to understand their specific characteristics (Hammerness *et al.*, 2005). The dissemination of the proposals for teaching approaches must consider their interaction with the teacher's PCK that is complex and relational and offer the teaching staff opportunities to think about the didactic materials to acquire complex and relational knowledge, based on deep comprehension of the topic and their capacity to adapt it to the students' needs (Clotfelter *et al.*, 2007).

Having reached this point, it must be remembered that in educational research, generalization of the research outcomes is not always based on generalizing a random population sample (statistical generalization); many research approaches point towards a generalization of a model presenting findings as specific cases of a more general model (Frick, 1998). A view is adopted that is oriented towards circumstantial evidence of the processes observed and that what happened is probably caused by the intervention (Maxwell, 2004). This focus of the design research benefits from pragmatic philosophy, that validates the theory depending on its capacity to explain phenomena and make changes in the world (Dewey, 1938; Wong and Pugh, 2001; and Juuti and Lavonen, 2006).

26.6 PERSPECTIVES OF DESIGN-BASED RESEARCH AND EVALUATION OF TLS

We started the chapter by claiming that the ultimate goal of physics education research should be to contribute to the improvement of teaching and learning processes in real-life contexts. However, the literature highlights the gap between research results and their impact on the classroom (Broekkamp *et al.*, 2007). The aim to reduce this gap and to promote research in specific contexts to provide new knowledge from the teaching-learning process and to support the general theories with specific outcomes has been a persistent concern since research began on teaching sciences in the 20th century. More than 100 years ago, the American philosopher J. Dewey, a major force in pragmatic philosophy, stated in his article in issue 1 of the *General Science Quarterly* journal (nowadays, *Science Education*) in 1916 (Dewey, 1916) and since reproduced on the request of the editor in 1945 (Dewey, 1945):

> "I can sum up by saying that it seems to me that our present methods too largely put the cart before the horse, and that when we become aware of this mistake, we are all too likely to cut the horse entirely loose from the cart and let him browse around at random in the pastures without getting anywhere. What we need is to hitch the horse of concrete experience with daily occupation and surroundings to a cart loaded with specialized scientific knowledge" (p. 8).

In addition, in 1999, S. Toulmin indicated that:

> "For the future, then, the key notion in any new theory of knowledge needs to be *practice*. In place of the *foundationalist* theories that held centre stage from Descartes to Russell, we shall do better to develop a new *praxeology* –the term is Kotarbinski's (1965)– that asks what

> procedures are efficacious in any given rational enterprise, on what conditions, and for what practical purposes" (p. 62).

We think that these quotes are still valid, although one of them is more than a century old. Henderson and Dancy (2009) show that many teachers of introductory physics courses are aware of Research-Based Instructional Strategies (RBIS) and are willing to try them, but they may be inappropriately implemented and/or abandoned. Efforts to disseminate RBIS at conferences and PER workshops have been successful in motivating and promoting teachers' interest in innovative strategies in physics teaching, but the high dropout rate suggests that teachers do not have the necessary knowledge to understand the fundamental features of a RBIS or that they underestimate the contextual factors that often present difficulties for innovative strategies. The DBR research line takes up this challenge and makes its main focus on the development of "humble theories" that provide knowledge about specific teaching-learning processes in a domain or field of the curriculum, and research-based teaching materials.

In this chapter, we have described and analyzed teaching proposals that attempt to give a detailed explanation of the teaching strategies that are implicitly covered and that present design and evaluation procedures that go beyond demonstrating the products for the classrooms and the learning achieved before and after the intervention. We have shown that RBD focuses primarily on the importance of implementation, usefulness and effectiveness of the designed product (TLS) in the classroom. DBR attaches particular importance to the involvement of teachers, learners and educational authorities, which brings PER closer to their needs. Intermediate frameworks have been described that were proposed by different teaching approaches and humble teaching-learning theories based on the general theories and the empirical studies in science teaching. We have gone into greater depth explaining the design process for educational interventions with the definition of the so-called design tools that aim to build bridges between the general theories and the specific intervention proposals in the classroom. Finally, we have demonstrated that one of the main challenges for these teaching proposals, that have demonstrated the improvement of learning in local contexts, is transferring this knowledge to other teachers and countries.

It is still early days in terms of having an agreement on what constitutes design-based research, why it is important and the methods to perform it. However, it has progressed, and we now know that this type of interventionist research in specific educational contexts requires detailed explanations of the implicit and explicit decisions that are taken regarding design and implementation.

The community of researchers and teachers should take on the challenge of grounding our novel lessons, activities, and instructional strategies for teaching specific physics contents in credible and useful research, while we contribute more generally to the methodological base to make progress on new theories regarding Teaching Physics. Over the next few years, as the TLS design-based research is conducted, we hope that the dialog can continue.

REFERENCES

Alghamdi, A. H. and Li, L., Int. J. Educ. Res. **1**(10), 1–12 (2013).
Ametller, J. *et al.*, Curric. J. **18**(4), 479–492 (2007).
Andhika, J. *et al.*, AIP Conf. Proc. **1708**, 070005 (2016).
Artigue, M., *Mathematics Didactics as a Scientific Discipline*, edited by R. Biehler *et al.* (Kluwer Academic, Dordrecht, 1992), pp. 7–39.
Artigue, M., *Encyclopedia of Mathematics Education*, edited by S. Lerman (Springer, New York, 2014), pp. 159–162.
Barab, S. A. and Squire, K. D., J. Learn. Sci. **13**(1), 1–14 (2004).
Barell, J. F., *Problem-based Learning: An Inquiry Approach* (Corwin Press, 2006).
Barniol, P. and Zavala, G., Phys. Rev. Spec. Top. Phys. Educ. Res. **10**(1), 010121 (2014).
Becerra-Labra, C. *et al.*, Int. J. Sci. Educ. **34**(8), 1235–1253 (2012).
Bell, P., Educ. Psychol. **39**(4), 243–253 (2004).
Broekkamp, H. and van Hout-Wolters, B., Educ. Res. Eval. **13**(3), 203–220 (2007).
Brookes, D. T. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020148 (2020).
Carr, W. and Kemmis, S., *Becoming Critical: Education Knowledge and Action Research* (Routledge, 1986).
Champagne, A. B. *et al.*, Am. J. Phys. **48**(12), 1074–1079 (1980).
Clement, J., Am. J. Phys. **50**(1), 66–71 (1982).
Clotfelter, C. T. *et al.*, Econ. Educ. Rev. **26**(6), 673–682 (2007).
Cobb, P. and Gravemeijer, K., *Handbook of Design Research Methods in Education: Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching*, edited by A. E. Kelly *et al.* (Routledge, 2008), pp. 68–95.
Cobb, P. *et al.*, Educ. Res. **32**(1), 9–13 (2003).
Coelho, R. L., Sci. Educ. **19**(1), 91 (2010).
Cohen, L. *et al.*, *Research Methods in Education* (Routledge, 2013).
The Design-Based Research Collective, Educ. Res. **32**, 5–12 (2003).
Dewey, J., Sci. Educ. **1**, 3–9 (1916).
Dewey, J., Sci. Educ. **29**, 119–123 (1945).
Dewey, J., *Experience and Education* (Macmillan, New York, 1938).
DiSessa, A. A., Cognit. Instruction **10**(2–3), 105–225 (1993).
Docktor, J. L. and Mestre, J. P., Phys. Rev. Spec. Top. Phys. Educ. Res. **10**(2), 020119 (2014).
Driver, R. *et al.*, *Children's Ideas in Science* (Oxford University Press, Milton Keynes, 1985).
Duit, R. (2009) "Bibliography—Students' and Teachers' Conceptions and Science Education." Resource document, see https://archiv.ipn.uni-kiel.de/stcse/.
Duit, R. *et al.*, *Developing Standards in Research on Science Education*, edited by H. E. Fischer (Taylor & Francis, London, 2005), pp. 1–9.
Duit, R. *et al.*, *Science Education Research and Practice in Europe: Retrospective and Prospective*, edited by D. Jorde and J. Dillon (Sense Publishers, 2012), pp. 13–37.
Duschl, R. A., *Improving Science Education—The Contribution of Research*, edited by R. Millar *et al.* (Open University Press, Buckingham, 2000).
Easterday, M. *et al.*, *Proceedings of International Conference of Learning Sciences*, edited by Polman *et al.* (International Society of the Learning Sciences, Boulder, CO, 2014), pp. 317–324.
Ellis, B. D., J. Hist. Ideas **23**(2), 273–278 (1962).
Etkina, E. *et al.*, J. Phys. Conf. Ser. **1882**(1), 012001 (2021).
Etkina, E. *et al.*, *Explore and Apply*, 2nd ed. (Pearson, San Francisco, CA, 2019a).
Etkina, E. *et al.*, *Explore and Apply*, 2nd ed. (Pearson, San Francisco, CA, 2019b).
Etkina, E. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **4**(2), 020108 (2008).
Etkina, E. *et al.*, Phys. Educ. Res. **5**, 010109 (2009).
Fensham, P., *Research in Science Education—Past, Present, and Future*, edited by H. Behrendt *et al.* (Kluwer Academic Publishers, Dordrecht, 2001), pp. 27–41.
Fraser, J. M. *et al.*, Rep. Prog. Phys. **77**(3), 032401 (2014).
Frick, R. W., Instrum. Comput. **30**, 527 (1998).
Galili, I. and Tseitlin, M., Sci. Educ. **12**(1), 45–73 (2003).
Gil Pérez, D. *et al.*, *Cómo Promover el Interés por la Cultura Científica? [How to Promote Interest on Scientific Culture?]* (UNESCO Santiago-Chile, 2005).
Gil, D. and Carrascosa, J., Sci. Educ. **78**(3), 301–315 (1994).
Gil, D. *et al.*, *La Enseñanza de las Ciencias en la Educación Secundaria [Science Education at Secondary School]* (Horsori, Barcelona: Horsori, 1991), p. 232.
Gil, D. *et al.*, Sci. Educ. **11**(6), 557–571 (2002).
Guisasola, C *et al.*, *Science Education in Focus*, edited by M. V. Thomase (Nova Science Publisher, 2008), pp. 55–85.
Guisasola, J. *et al.*, Phys. Rev. Phys. Educ. Res. **13**(2), 020139 (2017).
Guisasola, J. *et al.*, *Research and Innovation in Physics Education: Two Sides of the Same Coin* (Springer, Cham, 2020), pp. 109–118.

Guisasola, J. *et al.*, *Fundamental Physics and Physics Education Research* (Springer, Cham, 2021), pp. 163–174.

Guisasola, J. *et al.*, Revista de Enseñanza de la Física **31**(2), 57–69 (2019).

Hammerness, K. *et al.*, *Preparing Teachers for a Changing World*, edited by L. Darling-Hammond and J. Bransford (Jossey-Bass, San Francisco, 2005), pp. 358–389.

Hazelkorn, E. *et al.*, "Science education for responsible citizenship," Report to the European Commission of the expert group on science education (2015).

Henderson, C. and Dancy, M. H., Phys. Rev. Spec. Top. Phys. Educ. Res. **5**(2), 020107 (2009).

Heron, P. R. and McDermott, L. C., Opt. Photonics New **9**, 30 (1998).

Heron, P. R. L., *Proceedings of the Enrico Fermi Summer School on Physics Education Research*, edited by Redish and Vincentini (Italian Physical Society, Varenna 2003).

Heron, P. R. *et al.*, *Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and Mathematics (STEM) Education* (American Association for Advancement of Science, Washington,DC, 2005), pp. 33–37.

Hestenes, D. *et al.*, Phys. Teach. **30**(3), 141–158 (1992).

Hewitt, P. G., *Conceptual Physics* (Pearson Educación, 2002).

Hjalmarson, M. and Lesh, R., *Handbook of Design Research Methods in Education: Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching*, edited by A. E. Kelly *et al.* (Routledge, London, 2008), pp. 96–110.

Hodson, D., *International Handbook of Research in History, Philosophy and Science Teaching* (Springer, Dordrecht, 2014), pp. 911–970.

Hurd, P. D., Sci. Educ. **75**(6), 723–732 (1991).

Jiménez-Aleixandre, M. P. *et al.*, Sci. Educ. **84**(6), 757–792 (2000).

Juuti, K. and Lavonen, J., Nord. Stud. Sci. Educ. **2**(2), 54–68 (2006).

Kelly, A. E. *et al.*, *Handbook of Design Research Methods in Education: Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching* (Routledge, 2008), pp. 3–18.

Kortland, J. and Klaassen, C. J. W. M., *Proceedings of the Symposium in Honour of Piet Lijnse at the Time of his Retirement as Professor of Physics Didactics at Utrecht University* (CDBeta Press, 2010).

Kotarbinski, T., *Praxeology: An Introduction to the Sciences of Action* (Pergamon Press, New York, 1965).

Kuhn, T., ISIS **45**(276), 29–32 (1984).

Leach, J. and Scott, P., Stud. Sci. Educ. **38**, 115 (2002).

Leach, J. *et al.*, *Improving Teaching and Learning in Science: Towards Evidence-Based Practice*, edited by R. Millar *et al.* (RoutledgeFalmer, London, 2006), pp. 79–99.

Leach, J. *et al.*, *Designing Theory-Based Teaching-Learning Sequences for Science Education, Utrecht University, FIsme Series on Research in Science Education* edited by Kortland and Klaassen (CDBeta Press, 2010), p. 64.

Leslie-Pelecky, D. L., Phys. Teach. **38**(3), 165–167 (2000).

Lijnse, P. L. and Klaassen, C. W. J. M., Int. J. Sci. Educ. **26**(5), 537 (2004).

Lijnse, P. L., Sci. Educ. **79**(2), 189–199 (1995).

Lijnse, P. L., *Trends in European Research in Science Education. Second European Summerschool* (Tessaloniki, 1994).

Marquit, E., Am. J. Phys. **58**(9), 867–870 (1990).

Matthews, M. R., *International Handbook of Research in History, Philosophy and Science Teaching* (Springer, Dordrecht, 2014), pp. 1585–1635.

Maxwell, J. A., Educ. Res. **33**, 3 (2004).

McComas, W. F. *et al.*, *The Nature of Science in Science Education. Rationales and Strategies*, edited by W. F. McComas (Kluwer Academic Publishers, The Netherlands, 2000), pp. 3–39.

McDermott, L. C. and Shaffer, P. S., *Tutorials in Introductory Physica, Instructor's Guide* (Jhon Wiley & Sons. Inc., NY, 2003).

McDermott, L. C. and Shaffer, P. S., *Physics by Inquiry* (Jhon Wiley & Sons. Inc., NY, 1996).

McDermott, L. C., Am. J. Phys. **59**(4), 301–315 (1991).

McKenney, S. and Reeves, T. C., *Conducting Educational Design Research* (Routledge, 2018).

Meheut, M. and Psillos, D., Int. J. Sci. Educ. **26**, 515 (2004).

Mortimer, E. F. and Scott, P. H., *Meaning Making in Secondary Science Classrooms* (Open University Press, Maidenhead, 2003).

National Research Council, *A Framework for K-12 Science Education: Practices, Crosscutting Themes, and Core Ideas* (National Academies Press, Washington, DC, 2012).

National Research Council, *Guide to Implementing the Next Generation Science Standards. Committee on Guidance on Implementing the Next Generation Science Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education* (The National Academies Press, Washington, DC, 2015).

Nersseian, N. J., Sci. Educ. **4**, 203–226 (1995).

Nieveen, N., *An Introduction to Educational Design Research*, edited by T. Plomp and N. Nieveen (SLO, Enschede, 2009), pp. 89–101.

Osborne, R. and Wittrock, M., Sci. Educ. **67**, 489–508 (1983).

Piaget, J., *Psychology and Epistemology: Towards a Theory of Knowledge* (Grossman, New York, 1971).

Pintó, R., Sci. Educ. **89**(1), 1–12 (2005).

Posner, G. J. *et al.*, Sci. Educ. **66**(2), 211–227 (1982).

Psillos, D. and Kariotoglou, P., *Iterative Design of Teaching-Learning Sequences: Introducing the Science of Materials in European Schools* (Springer, 2015).

Reeves, T. C., *Educational Design Research*, edited by J. van Den Akker *et al.* (Routledge, 2006), pp. 52–66.

Rudge, D. and Howe, E., Sci. Teach. **71**(9), 52–57 (2004).
Ruthven, K. *et al.*, Educ. Res. **38**(5), 329–342 (2009).
Sandoval, W. A. and Bell, P., Educ. Psychol. **39**(4), 199–201 (2004).
Savall, F. *et al.*, Phys. Rev. Phys. Educ. Res. **15**(2), 020138 (2019).
Taber, K. S., *Science Education* (SensePublishers, Rotterdam, 2017), pp. 119–131.
Testa, I. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020146 (2020).
Tiberghien, A., *Research in Science Education in Europe: Current Issues and Themes* (The Falmer Press, London, 1996), pp. 100–114.
Tiberghien, A., *Improving Science Education—The Contribution of Research* (Open University Press, Buckingham, 2000), pp. 27–47.
Tiberghien, A. *et al.*, Int. J. Sci. Educ. **31**(17), 2275–2314 (2009).
Tobin, K. and Tippins, D. J., *The Practice of Constructivism in Science Education*, edited by K. Tobin (AAAS, Washington, 1993), pp. 3–21.
van den Akker, J., *Design Methodology and Developmental Research in Education and Training*, edited by J. van den Akker *et al.* (Kluwer Academic Publishers, The Netherlands, 1999), pp. 1–14.
Verdu, R. and Martinez-Torregrosa, J., Doctoral dissertation (University of Valencia, Spain, 2004), see http://rua.ua.es/dspace/handle/10045/2782.
Vosniadou, S., Learn. Instr. **4**, 45–69 (1994).
Wandersee, J. H., J. Res. Sci. Teach. **29**(4), 423–434 (1992).
Wieman, C. E., Proc. Natl. Acad. Sci. U.S.A. **111**(23), 8319–8320 (2014).
Wong, D. and Pugh, K., J. Res. Sci. Teach. **38**(3), 317–336 (2001).
Zuza, K. *et al.*, Phys. Rev. Spec. Top. Phys. Educ. Res. **10**(1), 010122 (2014).
Zuza, K. *et al.*, Phys. Rev. Phys. Educ. Res. **16**(2), 020110 (2020).

CHAPTER 27 EPILOGUE

Dean A. Zollman

Zollman, D. A., “Epilogue,” in *The International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, New York, 2023), pp. 27-1–27-8.

27.1 INTRODUCTION: A LITTLE HISTORY

The first publications in physics education research can be traced to the early 20th century (Meltzer and Otero, 2015). At that time, studies in physics education primarily focused on innovations or improvements in the teaching of physics. Some of these “experiments” were influenced by thinking in the broader area of education beyond physics or even STEM education.

Beginning in about the mid-20th century, a sense of awareness of the state of science, mathematics and technology education in the United States spawned by Sputnik led to a curriculum development movement in the sciences, particularly in physics. Several K–12 curriculum projects focused on active involvement of the students in the learning process and included a focus on reasoning and other skills as well as physics concepts. Two major high school physics curricula in the U.S., Physical Science Study Committee (PSSC, 1960) and Project Physics (Holton, 2003), had long lasting effects on the teaching of physics at both high school and college levels. This influence was very strong in the United States but also affected physics teaching in other parts of the world. At least equally important were the efforts in elementary school science. Of particular importance to the physics community was the Science Curriculum Improvement Study (SCIS) led by Robert Karplus beginning in 1962. Karplus, who was an accomplished research physicist, had become interested in the teaching and learning of physics (Karplus and Their, 1969). The SCIS curriculum (Karplus and Their, 1971) applied the intellectual development model that Jean Piaget developed to the learning of elementary science. An important underlying theoretical framework, constructivism, proposed that knowledge is constructed in the mind of the learners based on their experiences and that the role of instruction was to create experiences that facilitated active student learning, rather than passively transferring from the teacher to the student. All of these curriculum development efforts, at least to some extent, were based on the notion of active student learning. Although they were not necessarily based on rigorous research on learning a particular physics topic, they were nevertheless based on general principles derived from science education or cognitive research. In addition to helping students learn content, goals in all

https://doi.org/10.1063/9780735425514_027

The International Handbook of Physics Education Research: Special Topics, Mehmet Fatih Taşar and Paula R. L. Heron

of these efforts were to facilitate deeper understanding, develop scientific reasoning and understand the process of science. While the initial effort focused on students in the early years of education, the influence of the thinking about how to teach STEM and what skills in addition to the content of physics went through the educational system (Karplus and Karplus, 1970).

During this time period, efforts to include PER in physics departments at the university level were taking shape. For example, in response to receiving the 1961 Oersted Medal, Francis Sears stated, “the most important thing today is for those in power to recognize that research in teaching is as important as research in physics” (Sears, 1962). A few years later, Eric Rogers, in a footnote in his 1969 Oersted response paper, describes an experiment with his students. He found that students could not explain the meaning of Newton’s Second Law even though they were quite good at calculations using the Second Law (Rogers, 1969).

Several histories of physics education and PER in the United States are available (Beichner, 2009; Cummings, 2011; and Meltzer and Otero, 2015). I do not know of similar histories of PER elsewhere.

Of course, physics and physics education are international endeavors. Two organizations that emphasize research and development in physics education and with beginnings at about the same time were founded in Europe. In 1960, a meeting at the UNESCO headquarters in Paris, attended by 86 participants representing 28 different countries, resulted in a resolution asking the International Union of Pure and Applied Physics (IUPAP) to establish a commission on the teaching of physics. Among the goals of this commission is “The collection, evaluation, and coordination of information and the stimulation of experiments at all levels of physics education” (French, 1980). By the end of 1960, the International Commission on Physics Education (ICPE) was created. Today, ICPE is also called Commission 14 of IUPAP and sponsors frequent conferences which include significant components on PER throughout the world.

In 1966, the Groupe International de Recherche sur l’Enseignement de la Physique (GIREP) was founded. As its name implies, this organization was devoted to research in physics education. The first GIREP conference was held in Lausanne, Switzerland in 1967. Early conferences tended to focus on secondary school physics teaching but quickly expanded to include university-level efforts as well. Today, curricular developments and PER are of equal importance in all GIREP activities.

All these efforts focused on improving the teaching and learning of physics, but many were also concerned about other aspects of student development. In 1971, McKinnon and Renner put additional emphasis on learning beyond the content of physics with their seminal paper “Are colleges concerned about in intellectual development?” (McKinnon and Renner, 1971). This titular question raised the issue of what the goals of physics and other STEM education were and if we were helping students to acquire skills beyond some factual knowledge of physics and to build on their previous knowledge to develop a deeper understanding of the physics concepts. These goals naturally led to the inclusion of active learning strategies for much of the curricular development by physics education researchers. Section I of *The International Handbook of Physics Education Research* (referred to as

IHPER): Learning Physics provides examples of the strategies for most topics in physics, while Sec. III discusses strategies that have been developed to facilitate the acquisition of skills beyond knowledge of the physics content matter.

As PER developed during the 20th century, assessment of student learning of content and other skills became increasingly important. As described in detail in *IHPER: Teaching Physics*, Sec. IV, a large number of instruments, methods and approaches are now available to help the PER community understand what the students know.

While snapshots of students' knowledge are important, it is equally important to understand how students came to acquire that knowledge and what in their backgrounds helped or hindered in that process. Section II of *IHPER: Learning Physics* and Sec. II of *IHPER: Teaching Physics* describe some tools that the community uses to understand the students' learning process and learning environments build on what the students bring to the classroom and facilitate the process.

27.2 PER AS PROLOGUE TO THE 21st CENTURY

While some PER can be conducted just to gain knowledge about student learning, much of it is and will be driven by pressures and challenges related to educating our students. For physics and other STEM disciplines, these challenges have been documented in many places and are frequently called educating for 21st century skills (Leshner and Scherer, 2018; APLU and UCSU, 2019; Stehle and Peters-Burton, 2019; and Widya *et al.*, 2019). In addition to the skills needed for student success, the challenge of making physics attractive to a broader segment of the population remains an issue for the physics community.

There are many different ideas about these 21st century skills. However, a large overlap exists between the traditional goals of undergraduate physics instruction and the 21st century goals. For example, the Phys21 report has the following list (Heron and McNeil, 2016):

- Physics-specific knowledge: Learning goals for physics-specific knowledge include the ability to use fundamental concepts such as conservation laws to solve problems, and competency in applying basic laws of physics in diverse topic areas and applied contexts. They also include the ability to represent physics concepts in multiple ways and solve problems involving multiple topic areas and disciplines.
- Scientific and technical skills: Learning goals for scientific and technical skills include the ability to solve ill-posed problems through experiments, simulations, and analytical models, determine follow-on investigations, and identify resource needs. They also include competencies in instrumentation, software, coding, and data analytics.
- Communication skills: Learning goals for communication skills include the ability to communicate orally and in writing with audiences that have a wide range of backgrounds and needs.

- Professional and workplace skills: Learning goals for professional and workplace skills include the ability to work in diverse teams; obtain knowledge about relevant technology resources; demonstrate familiarity with workplace concepts such as project management, budgeting, quality assessment, and regulatory issues; demonstrate effective management of difficult situations (including classrooms); and demonstrate awareness of career opportunities for physics degree holders and effective practices for job seeking.

Looking more broadly at undergraduate education in general, two organizations representing primarily public universities in the United States stated some of the goals of 21st century university education and quoted from Joseph Aoun's recent book *Robot-Proof: Higher Education in the Age of Artificial Intelligence* (Aoun, 2017) that to "survive and thrive in an era of smart machines and the increasing automation" He argues the goal is "to continue to develop the skills that are uniquely human" such as "not only communication, critical thinking, leadership, and teamwork but also the core human elements of curiosity, empathy, creativity, cultural agility, and entrepreneurialism" (APLU and UCSU, 2019).

This report recognizes that major changes in higher education will be needed to implement this "robot proof" education. Again, quoting Aoun, the report emphasizes "the need to ground this transformation in learning science." (APLU and UCSU, 2019). Thus, the door is open for physics education research to help drive these changes.

One way to look at a possible role for PER, that is related to 21st century physics learning, is to compare the skills list above with the chapters in these books. Clearly, all of the chapters in *IHPER: Learning Physics*, Sec. II, are closely related to the first bullet above, physics specific knowledge. For many years now, physics education researchers have been investigating students' learning of specific physics topics and developing active learning techniques to improve that learning. Bao and Koenig (2019) in describing PER efforts for the 21st century refer to this type of research as promoting "discipline-specific deep learning." No doubt this type of research and development will continue.

In addition, physics specific knowledge as described here goes beyond just "knowing physics" and includes problem solving, applications and multiple representations. Chapters in *IHPER: Learning Physics*, Secs. II and III, address issues related to these goals and form a foundation for much research that is to come later. Part of the investigations will necessarily involve how to adjust the content to provide room for other 21st century skills. An important aspect will include integrating the learning of physics with other aspects ranging from creativity to entrepreneurship.

Chapter 18 in *IHPER: Learning Physics* and Chap. 9 in *IHPER: Teaching Physics* directly address some of the goals listed in the second bullet, scientific and technical skills. In addition, all of the chapters that describe research and development on using computation and technology (primarily in Sec. III of *IHPER: Learning Physics* and Secs. I and II of *IHPER: Teaching Physics*) are important foundational studies for future work in this area. However, we need to know more about how to approach student work on ill-defined problems. For example, how do I make a problem truly ill-defined so that the student cannot find the answer anywhere on the web, but, not frustrate students because I am asking them to do something that is too far removed from what they have not been asked to do in their previous academic careers? Furthermore, today's students will have a lot of experience with various types of simulations, software, and maybe even data analytics. Thus, they will bring to their study of simulation analysis, and experimentation in physics some preconceived idea about how to use these skills in physics. We need to learn more so that we can build on their previous learning when appropriate and facilitate change when necessary.

Teaching physics students about communication skills has frequently not been an explicit part of the curriculum. Certainly, our students have frequently been required to explain their experimental results in lab reports. However, the number of reports and sometimes the amount of prose has changed as we have increased the emphasis on working in collaborative groups which submit one report per group. PER that has increased the emphasis on communication skills is the realization that students can become proficient at solving algorithmic problems but have very limited conceptual understanding of the laws of physics underlying the algorithms. Trying to facilitate conceptual learning in students and helping them communicate that learning has been a valuable contribution of PER. At the same time, collecting data on conceptual understanding from large numbers of students has necessitated the use of multiple-choice assessments (see several of the chapters in *IHPER: Teaching Physics*, Sec. IV). While these tests may collect valid data for the studies under investigation, they do very little to help students learn the communication skills needed in the 21st century.

The switch during on-line learning and teaching in 2020 provided an opportunity to conduct research on using technology to communicate with our students and for students to communicate with us (Pagoto *et al.*, 2021; and Banks and Vergez, 2022). In recent years, communication methods have expanded greatly. The emergency nature of knowledge delivery during the pandemic probably limited the usefulness of this research in determining students' preferences for methods for communicating their ideas and reasoning. As the types of methods for communication continue to increase, physics students, like many others, are likely to become very comfortable using audio-visual methods of communicating. In addition, short communications such as Twitter and SMS are frequently preferred by students when communicating with their peers. However, when our students get to the workplace, they will need to be able to prepare a variety of types of documents to communicate their findings. Discovering how to take advantage of their expertise with TikTok to help them learn scientific communication could be a useful investigation.

Only some of the professional workplace skills mentioned in the fourth bullet are explicitly part of some physics curricula in schools or universities. Working in teams, particularly in instructional laboratory settings, is becoming a strategy that many instructors include. Likewise, argumentation and computational skills have been integrated in some instruction (see *IHPER: Learning Physics*, Sec. III). To include many of the other skills will require creativity on the physics faculty and collaboration with people in other disciplines. An important aspect will be to be able to build on our students' background and previous learning in physics and elsewhere and efficiently expand the their academic experience to prepare them for the 21st century workplace.

27.3 BEYOND THE SKILLS

Almost all of the discussions of 21st century STEM students, or even students in general, include statements about the increasing diversity of these students in terms of ethnicity, socio-economic status, academic background, and physical and mental challenges. Section I of *IHPER: Special Topics* provides significant insights into some of the PERs in this area. An additional recent investigation compares how academic background is related to what we frequently called "talent" (Walton and Wieman, 2022). While Sec. I of *IHPER: Special Topics* shows that a lot of significant research has been undertaken in recent years, this area of PER still has much to do to help physics instructors at all levels communicate better with the broad range of students who could be successful at physics.

27.4 21st CENTURY GRADUATE EDUCATION IN PER

The U.S. National Academy of Sciences has expressed similar types of goals for graduate STEM education. In addition, the Academy has added specific goals for dissertation research such as

- Students should seek opportunities to work in cross-disciplinary and cross-sector teams during their graduate education and via extracurricular activities and be incentivized by their departments and faculty advisers to do so.
- Graduate programs and faculty should encourage and facilitate the development of student teams within and across disciplines (Leshner and Scherer, 2018).

The Academy report also includes many goals related to career preparation. These goals will not be discussed here.

Graduate education in PER somewhat naturally fits with the two goals listed above. By its nature, PER involves more than one traditional discipline and can involve several. However, there have been times during the past century when some of us have not taken advantage of our colleagues in the other disciplines as much as we could. Instead, we would learn what we needed of another discipline to get a project done. For 21st century graduate education, this report and others are advocating that we and our students actively collaborate with researchers in many disciplines that can be of value in completing physics education research. When we do this, we and our students will gain knowledge that can be applied to PER and learn and operate in an academic cultural setting that is different from that of the basic sciences. Thus, our students will be better prepared to pursue careers that can be broadly multi-disciplinary and are likely to change during their lifetimes.

Working in teams also has long been a part of PER. Our students have worked with their peers and mentors to investigate many of the difficult issues of learning and teaching physics. This approach should continue to be effective and should continue to prepare our students for a highly collaborative workplace in academia or elsewhere.

While PER graduate education seems to fit well with some of the recommendations for 21st century STEM graduate education, as a community we will continue to need to be aware of the changes that will require modifications in how we are helping our graduate students prepare for their careers.

27.5 CONCLUSIONS

At this time, the educational community is facing major changes, some brought about by the changing landscape in which our students operate, others by forces such as budgetary over which we have little or no control. For the physics education community to deal with these issues in an efficient and effective way, it needs strong evidence-based information with which to base instruction, mentoring and advising. PER will provide that information through carefully constructed and executed investigations. 21st century PER will continue to focus on fundamental issues related to teaching and learning physics. As described in these volumes, those issues go beyond learning the content of physics and include a wide range of cognitive and social issues. Thus, PER has a firm foundation for this future. As we see the new challenges facing students and society, the range of issues which will be appropriate for the physics education research community will continue to expand. At the same time, physics education researchers will need to maintain close ties with the rest of the physics community in secondary schools, higher education and the broader work and social communities.

ACKNOWLEDGMENT

Part of Sec. 27.1 of this chapter is an edited version from an unpublished paper by Sanjay Rebello and me.

REFERENCES

Here, I have limited the number of references to those closely related to the discussion. Many more references could have been included. Instead, throughout this chapter, I referred the reader to sections in these volumes where the authors have provided a significant number of up-to-date references.

Aoun, J., *Robot-Proof: Higher Education in the Age of Artificial Intelligence*, 1st ed. (MIT Press, 2017).

APLU and UCSU, *Delivering 21st Century Skills* (Coalition of Urban Serving University and Association of Public & Land Grant Universities, 2019), see http://www.aplu.org/commissions/urban-serving-universities/usu-publications.

Banks, D. P. and Vergez, S. M., J. Microbiol. Biol. Educ. **23**(1), e00012–e00022 (2022).

Bao, L. and Koenig, K., Discipl. Interdiscipl. Sci. Educ. Res. **1**(1), 2 (2019).

Beichner, R., in *Getting Started in PER*, edited by C. Henderson and K. Harper (AAPT, College Park, 2009), Vol. 2; available at https://www.compadre.org/Repository/document/ServeFile.cfm?ID=8806&DocID=1147.

Cummings, K., *A Developmental History of Physics Education Research (The National Academies' Board on Science Education Commissioned Papers, p. 24)* (National Academy of Sciences, 2011), see https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_072580.pdf.

French, A. P., Contemp. Phys. **21**(4), 331–344 (1980), see https://web.phys.ksu.edu/icpe/info/history/frenhis.htm.

Heron, P. L. and McNeil, L., *Phys21: Preparing Physics Students for 21st Century Careers* (APS & AAPT, 2016), see https://www.compadre.org/JTUPP/report.cfm.
Holton, G., Sci. Educ. **12**(8), 779–786 (2003).
Karplus, E. and Karplus, R., School Sci. Math. **70**, 398–406 (1970).
Karplus, R. and Their, H. D., *A New Look at Elementary School Science* (Rand McNally, 1969).
Karplus, R. and Their, H. D., *SCIS: The Science Curriculum Improvement Study*, edited by E. Victor and M. S. Lerner (Macmillan, 1971).
Leshner, A. and Scherer, L., *Graduate STEM Education for the 21st Century* (National Academies Press, 2018).
McKinnon, J. W. and Renner, J. W., Am. J. Phys. **39**, 1047–1052 (1971).
Meltzer, D. E. and Otero, V. K., Am. J. Phys. **83**(5), 447–458 (2015).
Pagoto, S. *et al.*, PLoS One **16**(8), e0256213 (2021).
PSSC, *Physics* (Raytheon Education Co, 1960).
Rogers, E. M., Am. J. Phys. **37**(10), 954–962 (1969).
Sears, F. W., Am. J. Phys. **30**(6), 401–403 (1962).
Stehle, S. M. and Peters-Burton, E. E., Int. J. STEM Educ. **6**(1), 39 (2019).
Walton, D. and Wieman, C., APS News **31**(7), 8 (2022), see https://www.aps.org/publications/apsnews/202207/backpage.cfm.
Widya, A. *et al.*, J. Phys.: Conf. Ser. **1317**, 012208 (2019).

INDEX

acceleration, 7-16
ACER-framework, 18-21; *See also* mathematics in physics
Active Physics, 15-8
active teaching technique, 26-19; *See also* design tools
activity guide program, 26-6; *See also* Teaching Learning Sequences
ad hoc enforcement of human rights, 1-5, 1-6
admission practices, 4-3–4-4; *See also* equity in physics graduate education
advanced mathematics, 18-16; *See also* mathematics in physics
advanced placement (AP), 1-29
aesthetic appreciation, 7-18
African science education, 13-25–13-26; *See also* Landscape of Physics Education
Age of Revolution, 9-5
agglomerative hierarchical clustering (AHC), 24-25
aggregating categories, 5-21; *See also* equity in research design
aims and values, 7-15; *See also* physics, aims and values of
Akaike Information Criterion (AIC), 24-20
Algodoo, 21-17–21-18; *See also* microworlds
al-Ḥasan ibn al-Haytham, 22-2
Alhazen. *See* al-Ḥasan ibn al-Haytham
American Association of Physics Teachers (AAPT), 1-7, 15-7
American Institute of Physics (AIP), 13-17
American Library Association (ALA), 7-2
American Physical Society (APS), 4-7, 13-4
American Sign Language (ASL), 1-30
analytical derivation game. *See* epistemic games
Anthropocene, 6-20–6-21
antiessentialist, 3-3
anti-gap studies, 3-4; *See also* gap-gazing
a priori coding schemes, 25-18, 25-22–25-23; *See also* coding and analyzing data
areas of interest (AOIs), 20-19
Aristotelean era, 14-4
artifacts, 25-13; *See also* qualitative methods
artificial intelligence, 7-3; *See also* scientific literacy
  and data science, 7-12
astronomy, 10-12
  revolution, 7-16
atoms, 17-11–17-12
attention-deficit hyperactivity disorder (ADHD), 1-11, 1-17
augmented reality (AR), 17-19; *See also* visualization and mathematization
  MAGNA AR visualization tool, 21-12
  for teaching and learning, 17-19
  visualizing vector fields with, 21-12
automatic word analysis software, 16-26

Bachelard, G., 9-11
Bayesian Information Criterion (BIC), 24-20
beauty, 7-15
behavior, "circus-freak" effect of, 3-12
bias
  gender, 6-9–6-13
  race, 6-13–6-15
bifurcated surveys, 11-12; *See also* epistemic beliefs
big data, 5-15–5-16
Black, Indigenous, and Women of Color (BIWOC), 3-12
blending, 18-22; *See also* mathematics in physics
body theory, 3-6
Boltzmann distribution, 7-12

Bonferroni correction, 24-8; *See also* quantitative methods
bootstrapping, 24-7; *See also* quantitative methods
boundary work, 10-5; *See also* scientific method
Bridge Program (BP), 4-7
Brief Electricity and Magnetism Assessment (BEMA), 24-3

card game bridge, 10-1
causal demonstrations, 10-3; *See also* scientific method
causal effect studies, 23-4; *See also* knowledge accumulation
causality principle, 9-13
cause, 12-4
Certain knowledge, 11-9
chaos theory, 22-3
Children's Learning in Science (CLIS) project, 26-5; *See also* research-based teaching-learning sequences; Teaching Learning Sequences
"circus-freak" effect of behavior, 3-12
classical era, 14-4
Classical Test Theory (CTT), 24-9; *See also* quantitative methods
  test reliability, 24-11–24-12
  test validity, 24-10
climate change, 6-20–6-22
clockwork universe, 12-4
Cluster Analysis (CLA), 24-25; *See also* quantitative methods
  multiple algorithms for, 24-25
  in PER studies, 24-27
  scree plot for, 24-26
  student responses to FCI, 24-26
coding and analyzing data, 25-17; *See also* conceptual level of content; cultural level of content; epistemic level of content; qualitative methods; textbook content analysis
  a priori coding schemes, 25-18, 25-22–25-23
  categorizing interview responses, 25-20–25-22
  coding overview, 25-18
  coding schemes, 25-20–25-24
  coding system, 16-4
  coding transcripts, 25-18
  componential analysis, 25-19
  developing coding scheme, 25-18–25-19
  generative coding, 25-18–25-19, 25-23–25-24
  representational level of content, 16-6
  taxonomic domains, 25-19
  working with codes, 25-19–25-20
Cognitive; *See also* mathematics in physics
  blending theory, 18-22
  maturation of children, 11-1
  semantics framework, 18-22–18-23
  theories, 21-4
Colonialist and Eurocentric values, 8-3
Colorado Learning Attitudes about Science Survey (CLASS), 5-14, 11-10, 11-11, 24-3, 24-4, 25-9, 25-13
Colorado Learning Attitudes about Science Survey for Experimental Physics (E-CLASS), 24-3, 24-4, 25-13
Commission 14 of IUPAP. *See* International Commission on Physics Education
Committee on Data for Science and Technology (CODATA), 23-7
"communicative focus" tool, 26-6; *See also* design tools; Teaching Learning Sequences
communities of practice (CoP), 5-20
community, 15-31–15-32; *See also* equity in physics graduate education
  practices, 4-7
Comparative Fit Index (CFI), 24-20
complementary science, 10-14
composing, 12-22
compulsory motherhood, 8-18; *See also* Meitner, L.
computational physics, 10-12

concepts ontological categories, 19-7–19-8; *See also* physics equations

conceptual
  knowledge of physics, 19-3
  understanding, 6-12

conceptual blending, 19-11; *See also* mathematization of physics
  framework, 18-22

Conceptual Conflict Collaborate Group (CG), 12-26

Conceptual Evaluation of Electricity and Magnetism (CSEM), 24-2, 24-3

conceptual level of content, 16-4, 16-11; *See also* coding and analyzing data; textbook content analysis
  basic information, 16-11–16-12
  geographical distribution of affiliation of first author, 16-12
  incomplete definitions, 16-13–16-14
  inconsistent definitions, 16-13
  incorrect definitions, 16-14
  interdisciplinary concepts, 16-14–16-15
  validity of concepts, 16-13

Conceptual Physics, 15-8, 17-16

Conceptual Survey of Electricity and Magnetism (CSEM), 2-8, 2-15

confidence interval (CI), 24-11

Confirmatory Factor Analysis (CFA), 24-15

consensus views (CV), 9-19, 9-22; *See also* physics, aims and values of
  teaching NOS away from, 9-24

Consolidated Standards of Reporting Trials (CONSORT), 23-10

conspiracy theory, 9-2
  movements, 8-2

constant, 19-6; *See also* physics equations

constellation mentoring model, 4-7; *See also* equity in physics graduate education

constructivism, 13-5–13-6

constructivist practices, 13-6

content analysis (CA), 16-3, 16-24; *See also* qualitative methods; textbook content analysis
  content knowledge, 19-17
  content log, 25-16
  content specificity, 24-10
  ICT supported, 16-26–16-27
  of key techniques and methods, 16-24–16-27
  mixed methods with, 16-26
  theme-based quantitative, 16-25–16-26

context-based learning, 17-12

contextualized teaching methods, 13-1, 13-2; *See also* history and philosophy of science

contextualizing pedagogies. *See* contextualized teaching methods

conventionalism, 12-9; *See also* philosophy of physics

Coulomb's inverse square law, 10-14

Court of Justice of the European Union (CJEU), 10-1

Covid-19 pandemic, 8-2

critic, 9-8

Critical Race Theory (CRT), 3-5, 3-7, 13-4

critical race theory, 6-15

critique
  activity, 12-26
  of inductivism, 9-7–9-8

Cronbach's alpha, 24-11

Cultural History of Science (CHS), 9-25
  -mediated activities, 13-14

cultural level of content, 16-4, 16-19; *See also* coding and analyzing data; textbook content analysis
  basic information, 16-19–16-21
  ethical representations, 16-21–16-22
  gender representations, 16-21
  geographical distribution of affiliation of first author, 16-20

Culturally Engaging Campus Environment (CECE) model, 5-20
culture of a group, 4-3
curriculum; *See also* textbook and curriculum alignment
  Enacted Curriculum, 15-5
  four-level view of curriculum, 15-3
  making and enacting process, 15-1–15-2
  place and role of textbooks in, 15-3–15-5
  school curriculum, 15-1
  types, 15-2
curriculum-based research of textbooks, 15-9; *See also* textbook and curriculum alignment
  China, 15-11–15-12
  Ethiopia, 15-12–15-13
  Nigeria, 15-13–15-14
  physics textbook studies for single country, 15-11–15-18
  reviewed studies, 15-10–15-11
  Singapore, 15-14–15-15
  Slovenia, 15-15–15-16
  South Korea, 15-16–15-17
  Türkiye, 15-17–15-18

data processing, 25-16–25-17; *See also* coding and analyzing data; qualitative methods
data sources and collection, 5-15; *See also* equity in research design
  big data and existing data sets, 5-15–5-16
  considerations in data collection, 5-16–5-18
deductive approach, 25-2; *See also* qualitative methods
deficit research, 5-9
DeFT (Design, Functions, Tasks), 21-4
demarcation problem, 9-7, 10-1; *See also* methods and practices in physics
  boundary work, 10-5
  causal demonstrations, 10-3
  scientific and non-scientific, 10-4
  scientific knowledge, 10-3
  scientific method as rhetorical tool, 10-5
  scientific theories, 10-3
demonstrations, 1-15
descriptive analysis, 24-4–24-6; *See also* quantitative methods
design-based research (DBR), 26-2; *See also* research-based teaching-learning sequences
  design process, 26-10
  design solution, 26-11
  educational context, 26-10
  evaluation and redesign, 26-11–26-12
  gaps in, 26-8–26-9
  implementation, 26-11
  intermediate framework building, 26-3
  perspectives of, 26-22–26-23
  research methodology on teaching and learning, 26-9
  retrospective analysis, 26-11–26-12
  teaching experiments, 26-11
  teaching-learning goals, 26-10–26-11
  TLS efficacy, 26-12
design tools, 26-6; *See also* research-based teaching-learning sequences
  active teaching technique, 26-19
  for articulating intermediate frameworks, 26-13–26-21
  "communicative focus" tool, 26-6
  construction of TLS activities, 26-20
  epistemological elements and learning objectives, 26-15
  "epistemological-ontological content analysis" tool, 26-14–26-16
  explanatory development of Newton's laws, 26-16
  Learning Demand tool, 26-6, 26-16–26-17, 26-18
  Staging tools, 26-18
  students' conceptions of difficulty, 26-17
  TLS framework into design brief, 26-18

  tools for "staging a teaching sequence", 26-17–26-21
d'Espagnat, B., 12-12
Developmental Research, 26-6; *See also* Teaching Learning Sequences
Differential Item Functioning (DIF), 24-13
diffraction, 7-13
digital; *See also* intra-curricular processes; Physics Textbook Evaluation Literature; scientific literacy
  physics textbooks, 15-28–15-30
  textbooks, 17-21–17-24
  tools, 7-2–7-3, 21-3
  transformation, 7-5
dimensionality of object of research, 11-8–11-10; *See also* epistemic beliefs
  Certain knowledge, 11-9
  dimensions for epistemological beliefs, 11-9
  Simple Knowledge, 11-9
direct instruction, 1-15
disability, 1-1
  frameworks, 1-9
disabled learners, 1-1
  accessible laboratory tools, 1-13
  ad hoc enforcement of human rights, 1-5, 1-6
  analyzing research-based instructional strategies, 1-18–1-20
  articles written for and by practitioners, 1-11–1-17
  critiques of literature, 1-25–1-26
  demonstrations, 1-15
  direct instruction, 1-15
  disability frameworks, 1-9
  discussion, 1-25–1-29
  findings, 1-10
  findings for researchers, 1-17
  first-order barriers, 1-24
  funding programs, 1-9
  future directions, 1-29–1-31
  implications for practitioners, 1-26–1-28
  implications for researchers, 1-28–1-29
  instructional practices, 1-15–1-17
  laboratory setting, 1-11
  legal requirements, 1-4–1-5
  methods, 1-9, 1-13–1-14
  modifying existing equipment, 1-11–1-13
  second-order barrier, 1-23
  selection of articles, 1-9–1-10
  textbooks, 1-15
  UDL-aligned strategies, 1-19–1-20
  virtual simulations, 1-14
disabled people in STEM, 1-2–1-4, 1-5, 1-6
  employment and salary, 1-4
  instructors' experiences, 1-22–1-25
  population, 1-2–1-3
  students' experiences in postsecondary STEM, 1-20–1-22
  undergrad and graduate enrollment, 1-3–1-4
  underrepresentation of, 1-6–1-9
  understanding of, 1-17–1-18
discipline-based education research (DBER), 1-18, 5-6
discovery; *See also* theoretical practices and methods
  learning, 12-21
  myths, 10-19–10-20
displacement, 7-16
division of positions, 8-20–8-26; *See also* Meitner, L.
DO-IT program (disabilities, opportunities, internetworking, and technology), 1-5
domain, 11-17; *See also* epistemic beliefs
  -general and consensus-based view, 10-6
domain-specificity, 11-10; *See also* epistemic beliefs
  bifurcated surveys, 11-12
  epistemic beliefs, 11-13
  gender-specific student profiles, 11-12
  knowledge-production affected by, 11-11
  literature, 11-11
  physics-specific epistemic beliefs, 11-11

dominant mathematical model, 19-10–19-11;
    See also mathematization of physics
Duhem–Quine thesis, 10-17

Economic and Social Commission for Asian and the Pacific (ESCAP), 1-3
Educational Endowment Foundation (EEF), 23-3;
    See also knowledge accumulation
  Toolkit, 23-3–23-4
educational games, 21-18–21-19; *See also* visualization and mathematization
  Motion Mapper, 21-19–21-20
Educational Reconstruction, 26-7; *See also* Teaching Learning Sequences
educational research, 26-1
effect size, 24-7–24-8; *See also* quantitative methods; reporting practices
  reporting, 23-14
Einstein, A., 10-19–10-20, 22-3
  Bohr–Einstein debate, 12-4–12-5
  photoelectric effect, 12-23
Einstein was Here program, 13-10
electrical and magnetic experiments, 7-13
electric guitar, 7-14
electromagnetic induction (EMI), 15-28, 16-14
electronic sources of data, 25-14; *See also* qualitative methods
elementary geometrical elements, 18-15; *See also* mathematics in physics
emics of epistemic beliefs, 11-14; *See also* epistemic beliefs
  classes of representations, 11-16
  etic and emic approaches, 11-14
  mapping students' epistemological mindsets, 11-16
  p-prims, 11-14–11-15
  Reflective Judgment Model, 11-15
emotions evoked by natural phenomena, 7-17
empirical
  adequacy, 12-12
  -inductivists, 9-3
  propositions, 9-6
empiricism, 9-3
Enacted Curriculum (EC), 15-3, 15-5; *See also* curriculum
encyclopedic knowledge, 18-22
Enlightenment; *See also* philosophy of physics
  Age, 12-3
  entities, 12-21
  rationality, 8-13
Entlassungs–Zeugnis, 8-18
entropy, 7-6
epistemic beliefs (EB), 11-1, 11-2, 11-13, 11-22
  adequacy of measurement-tools, 11-22–11-23
  analysis of components, 11-14–11-16
  bifurcated surveys, 11-12
  Certain knowledge, 11-9
  classes of representations, 11-16
  common dimensions for, 11-9
  conceptual issues, 11-17–11-19
  domain, 11-17
  domain-specificity and, 11-10–11-14
  domain-specific literature, 11-11
  emics of, 11-14–11-16
  epistemic basing relation, 11-17–11-18
  epistemic cognition for Reflective Judgment Model, 11-15
  Epistemological Reflection Model, 11-19
  etic and emic approaches, 11-14
  evaluating knowledge claims and controversial issues, 11-4–11-7
  experimenter's regress, 11-19–11-22
  foundational issues, 11-16
  gender-specific student profiles, 11-12
  growing up to face relativism, 11-7–11-8
  hierarchies of skills, 11-3–11-4
  increasing dimensionality of object of research, 11-8–11-10

  interpretation of questionnaire-item, 11-18–11-19
  knowledge-production affected, 11-11
  mapping students' mindsets in understanding physics equations, 11-16
  from methodological questions to perceived congruence of field, 11-23–11-26
  personal epistemology, 11-18
  physics-specific epistemic beliefs, 11-11
  post-adolescent reasoning styles, 11-4
  p-prims, 11-14–11-15
  to proliferation, 11-2
  questionnaire items with critical questions, 11-24–11-25
  Reflective Judgment Interview, 11-5
  Reflective Judgment Model, 11-4, 11-21
  Reflective Reasoning, 11-6
  research on, 11-26–11-28
  responses of *Reflective Judgment Model*, 11-6
  Simple Knowledge, 11-9
  skill-theory, 11-3
Epistemic Beliefs Inventory (EBI), 11-10
epistemic games, 18-20–18-21, 19-12–19-13; *See also* mathematics in physics; mathematization of physics
epistemic knowledge, 7-11; *See also* scientific literacy
epistemic level of content, 16-4, 16-15; *See also* textbook content analysis
  basic information, 16-15
  experiments, 16-17
  geographical distribution of affiliation of first author, 16-16
  history and philosophy of science, 16-17–16-18
  nature of science, 16-18–16-19
  scientific literacy, 16-15–16-17
epistemological anarchism, 9-11
epistemological beliefs. *See* epistemic beliefs
Epistemological Beliefs Assessment for Physical Sciences (EBAPS), 11-10
epistemological framing, 19-12; *See also* mathematization of physics
epistemological-ontological content analysis tool, 26-14–26-16; *See also* design tools
epistemological pessimism in education research, 23-1–23-2; *See also* knowledge accumulation
Epistemological Reflection Model, 11-19
epistemology, 9-10–9-11
equilibrium or stationary distribution. *See* Boltzmann distribution
equity, 5-20–5-21
equity and inclusion, 2-1
  course-level equity in outcome, 2-8–2-10
  critiques, 2-20–2-24
  CSEM, 2-15
  equity in physics learning, 2-2, 2-4
  framework, 2-1–2-7
  future directions, 2-20–2-24
  gender and ethnicity, 2-1–2-2
  gender gap, 2-8
  HELPIEE framework, 2-4, 2-5, 2-6
  interventions, 2-18–2-20
  motivational beliefs, 2-11–2-15
  "my ability", 2-13
  path analysis part, 2-23
  perceived recognition, 2-12
  physics education equity research, 2-3
  physics identity, 2-12
  physics major-level equity in outcome, 2-7–2-8
  qualitative method research, 2-16–2-18
  self-efficacy, 2-11
  sense of belonging in physics courses, 2-12–2-13
  stereotype threats, 2-5, 2-15–2-16
  structural equation modeling, 2-22–2-24
  TEAM-UP report, 2-14
  underrepresented demographic student groups, 2-1
  values affirmation, 2-18–2-19
  verbal suggestions, 2-11

equity, diversity, and inclusion (EDI), 5-1
equity in physics graduate education, 4-1
  admission practices, 4-3–4-4
  community practices, 4-7
  constellation mentoring model, 4-7
  effects of marginalization, 4-4
  graduate assistantships, 4-7
  implications for researchers, 4-8–4-9
  mentoring, social, and academic support, 4-6–4-8
  microaggressions of racism and/or sexism, 4-6
  racialized imposter phenomenon, 4-6
  research on cultural and structural context, 4-3
  research on student experience, 4-5–4-8
  research on systemic and cultural factors as barriers to, 4-2–4-5
  student retention, 4-4–4-5
  work–life balance and mental health, 4-8
equity in research design, 5-1, 5-24
  addressing inequities, 5-7–5-8
  aggregating categories, 5-21
  base-level decisions, 5-2
  big data and existing data sets, 5-15–5-16
  challenges in recruiting research participants, 5-10–5-11
  changes to research design, 5-23–5-24
  conceptualizing research sample, 5-10–5-11
  considerations in collecting data, 5-16–5-18
  data sources and collection, 5-15
  DBER, 5-6
  deficit research, 5-9
  developing research goals and questions, 5-8–5-9
  discussion, 5-24–5-25
  initiating study on equity, 5-4–5-8
  inviting additional researchers, 5-6–5-7
  model minority myth, 5-21
  participants as decision making partners, 5-11–5-12
  PERBites, 5-23
  and physics education research, 5-1–5-2
  recruiting research participants, 5-10
  reflecting on results, 5-22
  research design, 5-1, 5-24
  researcher and team, 5-4, 5-5–5-6
  research instruments, 5-13–5-15
  research participants, 5-10
  safety and care for researchers and participants, 5-18–5-20
  sharing research findings, 5-22–5-23
  theoretical frameworks, data analysis, and interpretation, 5-20
  theories and definitions of equity, 5-20–5-21
errors, 24-8; *See also* quantitative methods
  estimation, 23-9
  type I error, 24-8
  type II error, 24-8–24-9
estimation error, 23-9; *See also* replication
ethnic-racial disparities, 8-2
ethnic/racial minority (ERM), 2-1
ethnography, 25-28–25-29; *See also* theoretical perspectives and frameworks
Ethnomathematics program, 8-12–8-13
etic approaches, 11-14
evidence-based active-engagement (EBAE), 2-9
evidence clearinghouses, 23-17; *See also* knowledge accumulation
experiential ethics of care, 8-13–8-14
experimental physics, 9-4
experimental practices and methods, 10-12; *See* also methods and practices in physics
  complementary science, 10-14
  Coulomb's inverse square law, 10-14
  Duhem–Quine thesis, 10-17
  experiment as mediator, 10-16
  experiment as mere hypothesis testing, 10-14–10-17
  experimentum crucis, 10-17
  experiment vs. theory, 10-15

  exploratory experiments, 10-15–10-16
  falsificationism, 10-14
  Goethe's theory of color, 10-16
  "great man history" phenomenon, 10-18, 10-20
  hypothetico-deductive method, 10-14
  inquiry-based science education, 10-16
  "myths" with respect to experiments, 10-13
  neglected role of assistants and technicians, 10-18
  serendipity, 10-18
  simplicity of experiments, 10-13–10-14
  theories to knowledge, 10-17
experimental science, 9-4
experimental *underdetermination*, 12-23
experimenter's regress, 11-19–11-22
experiments, 16-17
experimentum crucis, 10-17
explicitly anti-gap work, 3-3; *See also* gap-gazing
exploratory experiments, 10-15–10-16
exploratory factor analysis (EFA), 11-20, 24-15
eye-tracking studies, 20-19; *See also* graphs
  eye movements of experts and non-experts, 20-20
  fixation duration, 20-19
  TUG-K, 20-21

factor analysis, 24-15; *See also* quantitative methods
  confirmatory, 24-15
  exploratory, 24-15, 24-16–24-17
  PCA, 24-16
  results of EFA of first 15 items of FCI, 24-15, 24-16
  scree plot, 24-15, 24-16
  selection of number of factors, 24-15–24-16
falsifiability, 12-6
falsificationism, 10-14; *See also* experimental practices and methods
family resemblance approach (FRA), 9-22, 9-24
Faraday's law, 16-14
Features of Science (FOS), 9-21
feminist; *See also* Landscape of Physics Education
  criticism of science, 8-14–8-15
  and indigenous practices, 13-23
  and social justice pedagogies, 13-24
  standpoint theories, 6-15, 13-22–13-23
Feminist Theory Theater (FTT), 13-24
findings for researchers, 1-17
first-generation college students (FGCS), 2-10
first-order barriers, 1-24
Fisher, K., 11-3–11-4
5E-based learning environment, 19-18; *See also* mathematization of physics
Fleck, L., 9-11
Flesch Reading Ease (FRE), 16-23
Force and Motion Conceptual Evaluation (FMCE), 24-2, 24-3
Force Concept Inventory (FCI), 1-18, 2-8, 5-14, 6-5, 6-11, 11-12, 17-16, 20-14, 24-2, 24-3
  scree plot of first 15 items on, 24-16
  student responses to, 24-26
four-tier multiple-choice (4TMC), 15-14
4WADI (Four-tier Wave Diagnostic Instrument), 15-14
framework theory, 12-20
free-body diagrams, 21-1
freewriting, 12-26
functions and equations, 18-14–18-15; *See also* mathematics in physics
fundamental ontological-type questions, 12-15; *See also* philosophy of physics
fundamental quantities, 17-11
funding; *See also* knowledge accumulation
  agencies, 23-16
  programs, 1-9

gain scores, 24-6; *See also* quantitative methods
gamification, 21-18

gap-based studies, 3-4; *See also* gap-gazing
  advantages and limitations of, 3-4–3-5
  anti-gap studies, 3-4
gap-gazing, 3-2; *See also* physics education research
  avoiding, 3-2
  comparing among women, 3-4
  explicitly anti-gap work, 3-3
  gender-based studies, 3-2
  gender essentialism, 3-3
  need to include diverse voices, 3-4–3-5
  performance gaps, 3-3
  quantitative work, 3-2
  stereotypes of men and women, 3-3
gap research, 5-9
gender; *See also* epistemic beliefs; equity and inclusion; gap-gazing; Landscape of Physics Education
  bias, 6-9–6-13
  corollary, 6-19
  essentialism, 3-3
  and ethnicity/race, 2-1–2-2
  gaps in concept inventories, 2-8
  identity, 8-17
  inequalities, 8-3
  -science relationships, 13-21
  -specific student profiles, 11-12
gender and social justice, 8-1
  Colonialist and Eurocentric values, 8-3
  considerations, 8-26–8-28
  conspiracy theory movements, 8-2
  Covid-19 pandemic, 8-2
  differences in educational attainments, 8-4
  ethnic-racial disparities, 8-2
  gender inequalities, 8-3
  Meitner and SSI on physics teaching, 8-16–8-26
  socio scientific issues in science education, 8-5–8-10
  SSI and HP to promote social justice, 8-11–8-16
General Relativity Theory (GR), 16-19
generative coding, 25-18–25-19, 25-23–25-24; *See also* coding and analyzing data
geocentric theory, 7-16
GeoGebra, 21-15–21-16; *See also* user-created simulations
geometry, 18-15–18-16; *See also* mathematics in physics
  optics textbooks, 17-12
German physics society (DPG), 14-30
Global History (GH), 9-25
Goethe's theory of color, 10-16; *See also* experimental practices and methods
  color experimenting, 13-10
grade point average (GPA), 2-7
graduate assistantships, 4-7; *See also* equity in physics graduate education
Graduate Record Examination (GRE), 2-18, 4-3
grain size, 26-7–26-8; *See also* Teaching Learning Sequences
graphical analysis game. *See* epistemic games
graphs, 20-1, 20-21–20-22
  background, 20-1–20-4
  comprehension from literature, 20-2
  comprehension of diagrams, 20-4
  comprehension processes, 20-3
  construction of, 20-6–20-8, 20-10–20-12
  difficulties with graph construction, 20-11–20-12
  eye-tracking studies, 20-19–20-21
  graph slope, 20-17
  instruments for assessment, 20-12–20-16
  interplay of mathematics and physics, 20-16–20-19
  interpretation of graphs, 20-4–20-6, 20-8–20-10
  in mathematics education research, 20-4–20-8
  PER instruments, 20-15
  in physics education research, 20-8–20-12
  problems in studies on graphs, 20-18
  TCV-MP, 20-14
  TOGS, 20-15

  TUG-C, 20-13–20-14
  TUG-K, 20-12–20-13
  understanding of graphs, 20-5
  use of different representations, 20-14
"great man history" phenomenon, 10-18, 10-20
Groupe International de Recherche sur l'Enseignement de la Physique (GIREP), 27-2
guided problem solving (GPS), 26-18
Gymnasium, 8-17

Hake gain. *See* normalized gain
hard-of-hearing (HH), 1-30
harmonic oscillation, 10-22
Harvard Project of Physics (HPP), 9-15
heavy-weight philosophers, 12-8; *See also* philosophy of physics
hermeneutic circle, 12-25; *See also* philosophy of physics
  method, 12-24–12-26
Hierarchical Linear Modeling (HLM), 24-21–24-22; *See also* regression
hierarchical linear regression (HLR), 6-12
Higher Secondary School Certificate (HSSC), 16-24
historians of science, 9-14
historical "myth-conceptions", 10-9, 10-10; *See also* history of physics
historical experiments and instruments, 13-9–13-14; *See also* Landscape of Physics Education
historical-investigative approach (HI), 13-9, 13-12; *See also* Landscape of Physics Education
Historically Black Colleges and Universities (HBCUs), 4-7
historical-philosophical approaches in physics teaching, 9-15–9-23
history and philosophy of science (HPS), 10-8, 10-13, 13-1, 13-27, 15-13, 16-15, 16-17–16-18, 17-6; *See also* Landscape of Physics Education
  analytical framework, 16-18
  constructivism, 13-5–13-6
  grading, 13-8
  "greater diversity" in physics education, 13-4
  literature reviews, 13-8–13-9
  missing student perspective, 13-7
  nature of science, 13-6–13-7
  student voice, 13-7–13-8
  transformation into physicist, 13-4–13-5
  widely documented trends, 13-5–13-7
history of physics (HP), 8-3, 10-8
  addressing SSI using, 8-15–8-16
  Colonialist and Eurocentric values, 8-3
  considerations, 8-26–8-28
  conspiracy theory movements, 8-2
  Covid-19 pandemic, 8-2
  differences in educational attainments, 8-4
  Enlightenment rationality, 8-13
  ethnic-racial disparities, 8-2
  Ethnomathematics program, 8-12–8-13
  experiential ethics of care, 8-13–8-14
  feminist criticism of science, 8-14–8-15
  gender inequalities, 8-3
  historical "myth-conceptions", 10-9, 10-10
  historical framing and presentism, 10-9–10-11
  Meitner and SSI on physics teaching, 8-16–8-26
  misuse of, 10-9–10-11
  neglect of Islamic influences on, 10-20
  pessimistic meta-induction, 10-11
  recent developments, 10-11
  role of HP in physics classes, 8-19
  role of physics in establishing rationality, 8-13
  sexist deviation, 8-15
  socio scientific issues in science education, 8-5–8-10
  and SSI to promote social justice, 8-11–8-16
  use of, 10-11
  Whig-history, 10-9

history of science (HS), 8-12, 9-11, 9-23; *See also* physics, aims and values of
  as didactic tool, 9-16, 9-20
  goals, 9-17
  and NOS, 9-18, 9-23
  writing, 9-15
History, Philosophy, and Sociology of Science (HPSS), 9-2; *See also* physics, aims and values of
  humanization of science, 9-26
  replication process, 9-25
  role in physics teaching, 9-23–9-26
  teaching NOS away from CV, 9-24
Holistic Ecosystem for Learning Physics in an Inclusive and Equitable Environment (HELPIEE), 2-4, 2-5; *See also* equity and inclusion
  course-level equity in outcome, 2-8–2-15
  major-level equity in outcome, 2-7–2-8
  physics instructors, 2-6
Holocene, 6-20
Humanistic era, 14-4
humanization of science, 9-26
hypothetico-deductive (H-D) method, 10-14; *See also* experimental practices and methods

i3e (International Initiative for Impact Evaluation), 23-4
Ibn al-Haytham. *See* al-Ḥasan ibn al-Haytham
impartiality principle, 9-13
Inclusive Teaching Strategies Inventory (ITSI), 1-24
incommensurability, 12-20, 12-25
incomplete definitions, 16-13–16-14; *See also* textbook content analysis
inconsistent definitions, 16-13; *See also* textbook content analysis
incorrect definitions, 16-14; *See also* textbook content analysis
indigenous and academic domains, 13-25; *See also* Landscape of Physics Education
indigenous peoples, 13-21, 13-22; *See also* Landscape of Physics Education
inductive analysis, 25-2; *See also* qualitative methods
inductivism; *See also* physics, aims and values of
  critique of, 9-7–9-8
  empirical-inductivists, 9-3
  to logical positivism, 9-3–9-6
inertial mass, 26-16
inferential statistics, 24-6–24-7; *See also* quantitative methods
inquiry-based laboratory activities, 17-14
inquiry-based science education (IBSE), 10-16
Inquiry Level Index, 16-26
Institute of Education Sciences (IES), 23-16; *See also* knowledge accumulation
  research portfolio, 23-3
Institute of Physics (IoP), 1-6
Institutional Research Board (IRB), 5-18
Institutional Review Board (IRB), 25-7
institutional support of educational research, 23-3; *See also* knowledge accumulation
instructional practices, general, 1-15–1-17
instructional strategies, 12-19–12-26
  for developing critical mind-set, 12-22
instruction and curricula improvement, 12-18–12-19
instructors' experiences teaching disabled students, 1-22–1-25
instrumentalism, 12-4–12-5; *See also* philosophy of physics
Intended Curriculum (IC), 15-3
interactive historical vignettes (IHVs), 13-16
Interactive Online Laboratory system (iOLab), 21-9
internal force, 12-20
International Commission on Physics Education (ICPE), 27-2

International Handbook of Physics Education Research (IHPER), 27-2–27-3
International History, Philosophy, and Science Teaching (IHPST), 9-16
International Olympic Committee (IOC), 10-1
International Union of Pure and Applied Physics (IUPAP), 27-2
internet-based learning environments, 15-30
intersectionality, 3-5; *See also* physics education research
  intersectional work, 3-5–3-8
  matrix of domination, 3-5
Intersectionally Conscious Collaboration (ICC), 2-18
intervention research in physics education, 23-5–23-6; *See also* knowledge accumulation
interviews, 25-6; *See also* qualitative methods
  assessment validation, 25-9–25-10
  conducting and recording, 25-8
  exploring students' and faculty members' experiences, 25-11
  individual vs. focus group, 25-6
  protocol development, 25-7–25-8
  purposes of, 25-8–25-11
  teaching interviews, 25-12
  think-aloud vs. stimulated-recall vs. artifact-based interview, 25-7
  types of, 25-6–25-7
  understanding students' problem solving, 25-10–25-11
  unstructured vs. semi-structured vs. structured, 25-6–25-7
intraclass correlation (ICC), 14-9
intra-curricular processes, 15-26; *See also* textbook and curriculum alignment
  aspects of textbook-based physics learning, 15-26–15-27
  digital physics textbooks in, 15-28–15-30
  erroneous conceptual understanding, 15-27–15-28
  limitations of textbook treatments of acceleration, 15-27
Investigative Science Learning Environment (ISLE), 26-7; *See also* Teaching Learning Sequences
iOLab, 21-9; *See also* microcomputer-based laboratory
IPCC (Intergovernmental Panel on Climate Change), 9-1
Islamic science, 10-20
Item Response Theory (IRT), 24-12; *See also* quantitative methods
  models, 24-12
  multidimensional IRT, 24-13, 24-14
  1PL model, 24-14
  Rasch analysis, 24-14
  Rasch theory, 24-14–24-15
  3-parameter logistic model, 24-13
  2-parameter logistic model, 24-12
  unidimensional IRT, 24-13

JAWS (Job Access With Speech), 1-12

K–12 curriculum projects, 27-1
KAGOAD (Küresel Aynalarda Görüntünün Oluşumunu Anlatan Düzenek), 1-14
Kaiser Wilhelm Institute (KWI), 8-20
KAM-(Kolmogorov–Arnold–Moser)-Theorem, 18-9
Karlsruher Physikkurs (KPK), 14-30, 16-28
kinematic graphs, 21-1
kinematics concept test (KCT), 20-15
Kitchener and King (K&K), 11-7
knowledge, 9-13
  creation, 23-17
  and Social Imagery, 9-13
knowledge accumulation, 23-1; *See also* physics education research
  causal effect studies, 23-4

knowledge accumulation (*Continued*)
  criticisms, 23-1–23-6
  EEF Toolkit, 23-3–23-4
  epistemological pessimism in education research, 23-1–23-2
  evidence clearinghouses, 23-17
  funding agencies, 23-16
  IES research portfolio, 23-3
  incentives, 23-17
  institutionalizing practices for, 23-16–23-17
  institutional support of educational research, 23-3
  meta-analysis, 23-2
  preeminent journals, 23-16–23-17
  replication, 23-6–23-10
  reporting practices, 23-10–23-16
  state of intervention research, 23-5–23-6
  statistical approaches to synthesizing effects, 23-5–23-6
  trends in physics education, 23-4–23-5
knowledge claims and controversial issues, 11-4; *See also* epistemic beliefs
  post-adolescent reasoning styles, 11-4
  Reflective Judgment Interview, 11-5
  Reflective Judgment Model, 11-4
  Reflective Reasoning, 11-6
Knowledge in Pieces (KiP), 12-19; *See also* Teaching Learning Sequences
  framework, 26-4
Kuder–Richardson 20 (KR-20), 24-11

laboratory setting, 1-11
  accessible laboratory tools, 1-13
  methods and tools, 1-13–1-14
  modifying existing equipment, 1-11–1-13
Lagrange formalism, 18-2; *See also* mathematics in physics
Landscape of Physics Education, 13-2–13-3; *See also* history and philosophy of science
  African science education, 13-25–13-26
  CHS-mediated activities, 13-14
  contextualization through narratives, 13-14–13-20
  feminist and indigenous practices, 13-23
  feminist and social justice pedagogies, 13-24
  feminist standpoint theories, 13-22–13-23
  Feminist Theory Theater pedagogy, 13-24
  gender-science relationships, 13-21
  Goethe's historical color experimenting, 13-10
  historical experiments and instruments, 13-9–13-14
  historical-investigative approaches, 13-9, 13-12
  indigenous and academic domains, 13-25
  indigenous peoples, 13-21, 13-22
  narrative-focused methods, 13-14–13-20
  racism, 13-21, 13-22
  students experiment with historical electromagnetic instruments, 13-11
  tours, 13-9
  Ubuntu, 13-25–13-26
  women physicists trained in Muslim majority countries, 13-22
language shift, 12-7
law, 19-9; *See also* physics equations
Learned Curriculum (LC), 15-3
learner motivation, 12-16–12-17
learning
  modalities, 17-24
  skills, 17-8
Learning About STEM Student Outcomes (LASSO), 5-15
Learning Demand tool, 26-6, 26-16–26-17; *See also* design tools; Teaching Learning Sequences
  to analyze gap between students' difficulties & learning objectives, 26-18
lecture-based instruction(LB instruction), 2-9
Lexical bundles (LBs), 16-23, 16-26

LGBT (Lesbian, Gay, Bisexual, and Transgender), 3-1, 3-8
LGBTQ+, 4-2
LGBT research in STEM education, 3-8–3-13; *See also* physics education research
  "circus-freak" effect of behavior, 3-12
  LGBT persons in physics, 3-11–3-13
  LGBT persons in STEM, 3-9
  students, 3-9–3-10
  workforce, 3-10–3-11
light quanta, 12-23
linear regression, 24-18; *See also* regression
literacy, 7-1; *See also* scientific literacy
  data, 7-3
  DigComp, 7-4
  digital, 7-2
  ICT, 7-3
  Media and Information, 7-3
  types of, 7-4
Logic, 9-6
Logic of Scientific Discovery, *The*, 10-15
"logico-scientific" thinking, 12-17
logistic regression, 24-19; *See also* regression

machine learning, 24-28–24-29; *See also* quantitative methods
Madchengymnasium, 8-18
MAGNA AR visualization tool, 21-12; *See also* augmented reality
main ideas, 26-21
marginalization effect, 4-4; *See also* equity in physics graduate education
marginalized groups, 4-2
  research participants from, 5-10
Maryland Physics Expectation survey (MPEX), 11-10, 24-4, 25-9, 25-13
mass, 12-14, 12-15
matched sample, 25-3; *See also* qualitative methods
mathematics
  codes conceptual knowledge, 19-7
  elements, 18-13–18-17
  equations, 19-3–19-4
  model, 18-17–18-18
  procedures, 18-16–18-17
mathematics in physics, 18-1; *See also* mathematization of physics
  ACER-framework, 18-21
  advanced mathematics, 18-16
  application of, 18-5
  blending of frameworks, 18-24
  communication, 18-13
  complementarity of mathematics and physics, 18-4
  conceptual blending framework, 18-22
  educational perspective, 18-7–18-18
  epistemic games, 18-20–18-21
  exactness, precision and prediction, 18-11–18-12
  external frameworks, 18-21–18-24
  framework of cognitive semantics, 18-22–18-23
  frameworks for blending of physics and mathematics, 18-18–18-24
  functions and equations, 18-14–18-15
  geometry, 18-15–18-16
  historical perspective on interplay, 18-6–18-7
  idealization and approximation, 18-17
  Lagrange formalism, 18-2
  literature overview, 18-3
  mathematical model, 18-17–18-18
  mathematical procedures, 18-16–18-17
  mathematics in physics education, 18-13–18-17
  mathematization phase, 18-6–18-7
  models of mathematization, 18-17–18-18
  NoNiP-framework, 18-20
  outlook, 18-24–18-25
  perspectives on interplay, 18-2–18-7
  philosophical perspective, 18-4–18-5
  physical mechanism game, 18-20
  physical model, 18-17

mathematics in physics (*Continued*)
  physical objects, 18-1
  pictorial analysis game, 18-20
  post-mathematization phase, 18-7
  pre-mathematization phase, 18-6
  reduction of cognitive load, 18-11
  roles of, 18-10–18-13, 19-1
  scientific objectivity, 18-12–18-13
  specific frameworks, 18-19–18-21
  specific roles of, 18-11–18-13
  symbolic forms, 18-19–18-20
  symbols and notations, 18-13–18-14
  teachers and teaching, 18-8–18-10
  technical vs. structural dimensions, 18-10–18-11
  theoretical backup and algebraic expressions, 18-9
  vectors, 18-16
  Vergnaud's theory of conceptual fields, 18-23
mathematization, 21-2–21-3; *See also* mathematics in physics; visualization and mathematization
  degrees of, 21-6
  modeling cycle of physics, 21-7
  models, 18-17–18-18
  phase, 18-6–18-7
  theories in physics, 21-5–21-6
mathematization of physics, 19-1, 19-4, 19-10; *See also* physics equations
  areas of student difficulty, 19-13
  as blending, 19-11–19-12
  conceptual blending, 19-11
  dominant mathematical model, 19-10–19-11
  epistemic games, 19-12–19-13
  epistemological beliefs concerning, 19-12–19-13
  epistemological framing, 19-12
  through equations, 19-4
  5E-based learning environment, 19-18
  integrated approaches, 19-16
  as part of modeling, 19-10–19-11
  PCK model, 19-17–19-19
  physical reasoning, 19-10
  scientific modeling, 19-10
  students' comprehension of, 19-13–19-17
  studies concerning students, 19-14–19-15
  teachers' comprehension of, 19-17–19-19
Matilda effect, 10-20
matrix of domination, 3-5
Matthew-effect, 10-20, 10-21
Matthew–Matilda effect, 10-20–10-21
Maximum Likelihood Estimation (MLE), 24-20–24-21
maximum likelihood techniques, 24-20–24-21; *See also* regression
measurement of eye movement. *See* eye-tracking studies
measurement problem, 12-16
Media and Information literacy (MIL), 7-3
Meitner, L., 8-16, 10-20
  academic trajectory of, 8-16
  betrayal by work group, 8-23
  cis-hetero concept of compulsory motherhood, 8-18
  denial of prejudice, 8-17
  denunciation from Institute, 8-22
  division of positions, 8-20–8-26
  Entlassungs–Zeugnis, 8-18
  gender identity, 8-17
  impossibility of expanding contact networks, 8-20
  informal obstacle, 8-19–8-20
  in Kaiser Wilhelm Institute, 8-20–8-21
  Madchengymnasium, 8-18
  Meitner's academic trajectory, 8-16–8-26
  noble prize and atomic bomb case, 8-24–8-26
  political situation in Berlin, 8-21
  prohibitions, 8-19–8-20
  prohibition to access spaces, 8-20
  role of HP in physics classes, 8-19

  and SSI on physics teaching, 8-16–8-26
  studies on nuclear fission, 8-23–8-24
  two elements in standpoint feminism, 8-19
mentoring, 4-6–4-8; *See also* equity in physics graduate education
meta-analysis, 23-2; *See also* knowledge accumulation
meta-curricular frameworks, 15-19, 15-23, 15-24; *See also* textbook and curriculum alignment
  explanations, thought experiments and use of analogies, 15-23–15-24
  Finland and United States of America, 15-20–15-21
  Indonesia, 15-21–15-22
  international evaluations, 15-19
  nature of science in physics textbooks, 15-20–15-23
  science literacy, 15-19, 15-20
  scientific explanations, 15-23
  social issues and 21st century skills, 15-24–15-25
  South Korea, 15-22
  21st century skills, 15-25–15-26
  thought experiments, 15-23–15-24
  United States of America, 15-22
  Vietnam, 15-22–15-23
Methodology of Scientific Research Programs, 9-10
methods and practices in physics, 10-1, 10-2
  card game bridge, 10-1
  demarcation problem, 10-1, 10-3–10-5
  experimental practices and methods, 10-12–10-18
  future research, 10-23–10-24
  history of physics in physics teaching, 10-8–10-11
  nature of science, 10-5–10-8
  scientific inquiry, 10-5–10-8
  scientific method, 10-1–10-5
  theoretical practices and methods, 10-19–10-23
microaggressions of racism and/or sexism, 4-6; *See also* equity in physics graduate education
microcomputer, 21-8
microcomputer-based laboratory (MBL), 1-12, 19-19; *See also* visualization and mathematization
  iOLab, 21-9
  RealTime Physics, 21-8, 21-9
  Workshop Physics, 21-8
microworlds, 21-16–21-17; *See also* visualization and mathematization
  Algodoo, 21-17–21-18
minority serving institutions (MSI), 2-21
mixed reality (MR), 21-10–21-12
model, 10-21; *See also* theoretical practices and methods
  minority myth, 5-21
  and modeling, 10-21–10-23
  physical, 18-17
  in qualitative research, 25-2
Model Based View (MBV), 10-22
modeling instruction (MI), 2-9
modern
  era, 14-4
  physics, 6-16–6-20
  science, 8-27
modes of thought, 8-9–8-10
Modified Module Analysis (MMA), 24-28
Modified Module Analysis Partial (MMA-P), 24-28
module analysis, 24-27–24-28; *See also* network analysis
module analysis for multiple choice responses (MAMCR), 24-28
Motion Mapper, 21-19–21-20; *See also* educational games
motivational beliefs, 2-11; *See also* equity and inclusion
  gender difference in students', 2-14
  "my ability", 2-13

motivational beliefs (*Continued*)
  perceived recognition, 2-12
  physics identity, 2-12
  self-efficacy, 2-11
  students' sense of belonging in physics courses, 2-12–2-13
  verbal suggestions from others, 2-11
Multidimensional Item Response Theory (MIRT), 24-3, 24-13, 24-14
multimedia-based learning (MBL), 15-29
multimedia learning modules (MLM), 15-29
Muslim majority (MM), 13-22, 22-9
"my ability", 2-13

nano-art, 7-13
narrative-focused methods, 13-14; *See also* Landscape of Physics Education
  case histories, 13-16
  contextualized narratives, 13-16
  interactive historical vignettes, 13-16
  learners' engagement with stories, 13-17
  Project Physics, 13-17
  stories, 13-15
  storyboards by schoolchildren, 13-18
  storytelling by contemporary scientists, 13-19–13-20
  Tycho Brahe's model, 13-19
National Assessment of Education Progress (NAEP), 6-14
National Center for Educational Evaluation and Regional Assistance (NCEE), 23-3
National Center for Education Research (NCER), 23-3
National Center for Education Statistics (NCES), 23-3
National Center for Special Education Research (NCSER), 23-3
National Institutes of Health (NIH), 23-3
National Research Council provided Next Generation Science Standards, 6-2
National Science Education Standards (NSES), 15-21
National Science Foundation (NSF), 1-4, 23-16
natural language processing (NLP), 16-9
natural philosophy, 9-14
nature of science (NOS), 7-14, 8-5, 9-1, 10-3, 10-5, 10-6, 10-23, 12-21, 13-2, 13-6–13-7, 17-6, 26-14; *See also* methods and practices in physics; physics, aims and values of
  alternative approaches to scientific inquiry and, 10-7–10-8
  consensus-based approach to scientific inquiry and, 10-6–10-7
  construction of consensual views of, 9-15–9-23
  contents in physics textbooks, 16-18–16-19
  diverse NOS methods, 13-6
  domain-general and consensus-based view, 10-6
  in physics textbooks, 15-20–15-23
  scientific literacy, 10-5
  teaching NOS away from CV, 9-24
Nature of Scientific Inquiry (NOSI), 10-6, 10-23
Nature of Scientific Knowledge (NOSK), 10-6
network analysis, 24-27; *See also* quantitative methods
  module analysis, 24-27–24-28
  social, 24-27
Newton, I., 8-6, 9-3, 10-19–10-20, 12-21–12-22, 12-23
Newtonian physics, 6-5, 12-9
Newton's laws, 8-6, 26-16; *See also* physics equations
  second law, 10-22, 19-4–19-5
Next Generation Science Standards (NGSS), 6-5, 6-7
Nominal Response Model (NRM), 24-14
non-continuity of Circle, 9-6
NoNiP-framework, 18-20; *See also* mathematics in physics
normalized gain, 24-6
nuclear theory, social values in, 6-17–6-18
null hypothesis significance testing (NHST), 24-9

object, depiction of, 7-13
objectivity, 9-14, 9-15
observations, 25-14–25-16; *See also* qualitative methods
OECD (Organisation for Economic Co-operation and Development), 7-1
Oldenburg Epistemic Beliefs Questionnaire (OLEQ), 11-10
1PL model, 24-14
online resources, 17-23
  education, 17-2–17-3

parameter, 19-6; *See also* physics equations
Particle Data Group, 23-7
pedagogical content knowledge (PCK), 18-8, 19-17; *See also* mathematization of physics
  model, 19-17–19-19
pedagogical knowledge (PK), 19-17
pedagogy, 17-13; *See also* Physics Textbook Evaluation Literature
  analogies, 17-16
  historical aspects of textbooks & inclusion of laboratory work, 17-13–17-16
  historical development of physics education in U. S. A., 17-13
  inquiry-based laboratory activities, 17-14
PERBites, 5-23
perceived recognition, 2-12
performance gaps, 3-3; *See also* gap-gazing
Perry, W. G., 11-7–11-8
pessimistic meta-induction, 10-11; *See also* history of physics
phenomenography, 25-28; *See also* theoretical perspectives and frameworks
phenomenological primitives (“p-prims”), 12-19
philosophers of science, 12-6; *See also* philosophy of physics
philosophical hermeneutics, 12-25; *See also* philosophy of physics
philosophy
  by members of circle, 9-6
  post-positivist philosophy, 9-7
  of Science of 20th century, 9-7
  and six “instincts”, 9-5
philosophy of physics (PoP), 12-1, 12-9–12-13, 12-26–12-27
  cause, 12-4
  clockwork universe, 12-4
  composing, 12-22
  conventionalism, 12-9
  critique activity, 12-26
  discovery learning, 12-21
  distribution of students’ views on NOS, 12-24
  empirically adequate, 12-12
  Enlightenment Age, 12-3
  entities, 12-21
  experimental *underdetermination*, 12-23
  falsifiability and *scientific revolution*, 12-6
  framework theory, 12-20
  freewriting, 12-26
  fundamental ontological-type questions, 12-15
  heavy-weight philosophers, 12-8
  hermeneutic circle, 12-25
  for improving instruction and curricula, 12-18–12-19
  incommensurability, 12-20, 12-25
  instructional strategies for developing critical mind-set, 12-22
  internal force, 12-20
  language shift, 12-7
  learner motivation and interests, 12-16–12-17
  learning theories and instructional strategies, 12-19–12-26
  light quanta, 12-23
  “logico-scientific” thinking, 12-17
  mass, 12-14, 12-15
  measurement problem, 12-16
  nature of science, 12-21

philosophy of physics (PoP) (*Continued*)
  Newton, 12-21–12-22, 12-23
  Newtonian physics, 12-9
  and PER, 12-19–12-26
  philosophers of science, 12-6
  philosophers of science and collaborative groups, 12-22–12-24
  philosophical hermeneutics, 12-25
  photoelectric effect, 12-23
  physics and philosophy, 12-3–12-5, 12-5–12-9
  physics, philosophy, and field of education, 12-13
  P-prims, 12-19
  quantum mechanics, 12-7
  realism and *instrumentalism*, 12-4–12-5
  reflective writing and hermeneutical circle method, 12-24–12-26
  reforming textbook pedagogy and instruction, 12-13–12-16
  Romantic science, 12-9
  scientific controversies, 12-5
  student epistemology, 12-19–12-22
  textbooks, 12-2, 12-3
  theory and hypotheses, 12-21–12-22
  Theory–Theory, 12-19
  Titius–Bode law, 12-21
  wave-particle duality, 12-23
  Young and Fresnel wave theories, 12-23
photoelectric effect, 12-23
physical mechanism game, 18-20; *See also* mathematics in physics
physical model, 18-17; *See also* mathematics in physics
physical objects, 18-1; *See also* mathematics in physics
Physical Science Study Committee (PSSC), 13-17, 17-13, 27-1
physicists
  engagement in cultural environment, 6-17
  Japanese vs. American physicist, 6-19
physics, 7-2, 8-1, 9-2–9-3, 13-1; *See also* epistemic beliefs
  classroom, 6-3–6-9
  demographic makeup of physics discipline, 6-3
  early physics education research, 6-3–6-9
  education, 22-2
  establishing rationality, 8-13
  future classroom, 6-20–6-22
  identity, 2-12
  pedagogical histories and anthropology of, 6-16–6-20
  -specific epistemic beliefs, 11-11
  -specific knowledge, 27-4
  teacher education, 11-1
physics, aims and values of, 9-1
  Age of Revolution, 9-5
  Chair of History and Theory of Inductive Sciences, 9-5–9-6
  coherent, 9-9
  consensus views, 9-19, 9-22
  conspiracy theories, 9-2
  critic, 9-8
  critique of inductivism, 9-7–9-8
  demarcation problem, 9-7
  denying scientific results, 9-1
  empirical-inductivists, 9-3
  empirical propositions, 9-6
  empiricism, 9-3
  epistemological anarchism, 9-11
  epistemology, 9-10–9-11
  experimental, empirical, or inductive science, 9-4
  experimental physics, 9-4
  historical review of, 9-3
  inductivism and empiricism, 9-3
  from inductivism to logical positivism, 9-3–9-6
  Logic, 9-6
  Methodology of Scientific Research Programs, 9-10
  non-continuity of Circle, 9-6

  philosophies by members of circle, 9-6
  philosophy and six “instincts”, 9-5
  Philosophy of Science of 20th century, 9-7
  Popper x Kuhn debate, 9-10
  positivism, 9-5
  positivist epistemology, 9-5
  post-positivist philosophy, 9-7
  problem of scientific change and historicist turn, 9-7–9-11
  rational reconstruction, 9-11
  relativism, 9-21–9-22
  Research Programs, 9-10–9-11
  role of HPSS in physics teaching, 9-23–9-26
  science and knowledge, 9-4
  science studies, 9-12–9-15
  scientific results and fundamental right, 9-2
  Structure of Scientific Revolutions, The, 9-9, 9-10
  theoretical entities, 9-20
  thought style, 9-11
  Three-State Law, 9-5
  Vienna Circle, 9-6
  views of nature of science, 9-15–9-23
  V-NOS questionnaire, 9-19
physics as human endeavor, 6-1
  Anthropocene, 6-20–6-21
  climate change and future physics classroom, 6-20–6-22
  cold war and modern physics discipline, 6-16–6-20
  combatting bias, 6-9–6-15
  comparative anthropological study of high energy physics, 6-18–6-19
  conceptual understanding, 6-12
  critical race theory, 6-15
  demographic makeup of physics discipline, 6-3
  early physics education research, 6-3–6-9
  epistemological features of science, 6-8
  feminist standpoint theories, 6-15
  Force Concept Inventory, 6-5, 6-11
  gender bias, 6-9–6-13
  gender corollary, 6-19
  Holocene, 6-20
  Japanese vs. American physicist, 6-19
  NGSS, 6-5, 6-7
  pedagogical histories and anthropology of physics, 6-16–6-20
  physicists engagement in cultural environment, 6-17
  physics classroom, 6-3–6-9
  race bias, 6-13–6-15
  Science and Engineering Practices, 6-6
  social values in development of nuclear theory, 6-17–6-18
  Traweek’s ethnographic work, 6-19
  Warwick’s study of training in mathematics, 6-18
physics education research (PER), 1-9, 3-1, 4-2, 8-2, 9-2, 13-7, 15-6, 21-1, 22-1, 22-3, 22-4, 22-10–22-11, 24-1; *See also* knowledge accumulation; qualitative methods; quantitative methods
  alignment of physics textbooks and curricula in, 15-5–15-9
  avoiding gap-gazing, 3-2–3-5
  beginnings of, 22-4–22-7
  big data in PER, 5-15–5-16
  causal effect studies, 23-4
  communication methods, 27-5
  developmental history of visualization tools in, 21-20
  equity in research design, 5-1
  first publications in, 27-1
  gap-gazing, 3-2
  gender-based studies in, 3-2
  goals and methods of, 22-8
  histories of physics education and, 27-2
  instruments, 20-15
  intersectionality, 3-5–3-8
  LGBT research in STEM education, 3-8–3-13

physics education research (PER) (*Continued*)
  literature, 23-5
  parallel trends in physics education, 23-4–23-5
  performance gaps, 3-3
  Phys21 report list, 27-3–27-4
  as prologue to 21st century, 27-3–27-6
  quantitative work, 3-2
  research on equity in, 4-1–4-9
  research on gender, intersectionality, and LGBT+ persons in, 3-1–3-13
  sampling of recent work, 22-7–22-9
  state of intervention research in physics education, 23-5–23-6
  study of general learning skills, 22-5
  study of problem-solving strategies, 22-5
  teaching about communication skills, 27-5
  textbook mistakes, 17-10
  21st century graduate education in PER, 27-6–27-7
  unitary ontologies, 22-7
Physics Education Technology (PhET), 1-14; *See also* simulations
  simulations, 21-13–21-14
physics equations, 19-1, 19-19–19-20; *See also* mathematization of physics
  concepts ontological categories, 19-7–19-8
  constant, 19-6
  epistemological beliefs concerning mathematization, 19-12–19-13
  interpretation of, 19-6–19-7
  law, 19-9
  mathematical structure of, 19-6
  mathematics codes conceptual knowledge, 19-7
  mathematics equations, 19-3–19-4
  mathematization as blending, 19-11–19-12
  mathematization as part of modeling, 19-10–19-11
  meanings, 19-7–19-9
  Newton's second law, 19-4–19-5
  parameter, 19-6
  physics and mathematics, 19-2, 19-9–19-13
  in physics education, 19-3
  principle, 19-9
  property, 19-7
  role and status of scientific knowledge, 19-8–19-9
  students' comprehension of mathematization, 19-13–19-17
  system, 19-7
  teachers' comprehension of mathematization, 19-17–19-19
  verbalization of, 19-4–19-7
Physics Textbook Evaluation Literature (PTEL), 17-3, 17-25; *See also* textbooks, evaluation of physics
  analogies, 17-16
  content, 17-6–17-13
  development of, 17-5
  digital textbooks, 17-21–17-24
  distribution over time, 17-5
  gaps in, 17-26
  gender balance, 17-7–17-8
  geometric optics textbooks, 17-12
  historical aspects of textbooks and laboratory work, 17-13–17-16
  history and philosophy of science, 17-6
  inquiry-based laboratory activities, 17-14
  mistakes in textbooks, 17-10–17-12
  nature of science, 17-6
  online resources, 17-23
  pedagogy, 17-13–17-16
  physics education development in U. S. A., 17-13
  physics in context, 17-12–17-13
  presentation, 17-16–17-19
  printed textbooks, 17-20–17-21
  readability, 17-12
  scientific literacy, 17-8–17-10
  technology, 17-6–17-7

  text and visualizations, 17-18
  virtual and augmented reality technologies, 17-19
  visualization and conceptualization, 17-16–17-19
physics textbook expectations research (PTER), 14-2; *See also* textbook
  countries of PTER articles, 14-9–14-10
physics textbooks, 16-1. *See* textbooks
Piaget's genetic epistemology, 26-4; *See also* Teaching Learning Sequences
pictorial analysis game, 18-20; *See also* mathematics in physics
PISA (Programme for International Student Assessment), 7-2, 7-11, 17-10
  era, 14-4
  scientific literacy assessment framework for, 7-12
pluralism, 8-10
Popper, K., 10-15
Popper x Kuhn debate, 9-10
positivism, 9-5; *See also* physics, aims and values of
  inductivism to logical, 9-3–9-6
  positivist epistemology, 9-5
Post-Covid-19 era, 14-4
post-positivist philosophy, 9-7
Post-Sputnik era, 14-4
Potentially Implemented Curriculum (PIC), 15-3
p-prims, 11-14–11-15, 12-19
preeminent journals, 23-16–23-17; *See also* knowledge accumulation
preparation for future learning (PFL), 20-16
presentation, 17-16; *See also* Physics Textbook Evaluation Literature
  text and visualizations, 17-18
  virtual and augmented reality technologies, 17-19
  visualization and conceptualization, 17-16–17-19
presentism, 10-9–10-11; *See also* history of physics
Principal Component Analysis (PCA), 24-16
principle, 19-9; *See also* physics equations
printed textbooks, 17-20–17-21; *See also* Physics Textbook Evaluation Literature
probeware tools. *See* microcomputer-based laboratory
procedural knowledge, 7-11; *See also* scientific literacy
programmatic principles, 9-13
programming. *See* user-created simulations
Project Physics, 13-17
property, 19-7; *See also* physics equations
publication bias, 23-15; *See also* reporting practices

qualitative methods, 25-1, 25-30; *See also* physics education research
  artifacts, 25-13
  coding and analyzing data, 25-17–25-24
  collecting data, 25-5–25-6
  content log, 25-16
  deductive approach, 25-2
  electronic sources of data, 25-14
  gaining confidence in claims, 25-30
  inductive analysis, 25-2
  interviews, 25-6–25-11
  matched sample, 25-3
  model and theory, 25-2
  observations, 25-14–25-16
  processing data, 25-16–25-17
  reliability, 25-30
  research questions and study design, 25-3–25-5
  surveys, 25-12–25-13
  teaching interviews, 25-12
  theoretical perspectives and frameworks, 25-24–25-29
  validity, 25-29
quantitative form, 26-16
quantitative methods, 24-1, 24-29; *See also* physics education research
  Bonferroni correction, 24-8

quantitative methods (*Continued*)
  bootstrapping, 24-7
  categories, 24-1
  classical test theory, 24-9–24-12
  cluster analysis, 24-25–24-27
  descriptive analysis, 24-4–24-6
  effect sizes, 24-7–24-8
  factor analysis, 24-15–24-18
  gain scores, 24-6
  inferential statistics, 24-6–24-7
  item response theory, 24-12–24-15
  machine learning, 24-28–24-29
  network analysis, 24-27–24-28
  regression, 24-18–24-22
  research-based conceptual instruments, 24-2–24-4
  beyond significance testing, 24-9
  statistical methods, 24-4–24-9
  statistical power, 24-8–24-9
  structural equation modeling, 24-22–24-25
  type I error, 24-8
  type II error, 24-8–24-9
quantitative research, 24-1; *See also* equity and inclusion
  on course-level equity in outcome, 2-8–2-10
  motivational beliefs, 2-11–2-15
  on physics major-level equity, 2-7–2-8
quantum mechanics (QM), 7-18, 10-17, 12-7, 12-11, 22-3
quasi-history. *See* Whig-history

race bias, 6-13–6-15
racialized imposter phenomenon, 4-6
racism, 13-21, 13-22; *See also* Landscape of Physics Education
  microaggressions of, 4-6
radio broadcasting, 7-14
Rasch model. *See* 1PL model
rational reconstruction, 9-11
readability, 17-12
realism, 12-4–12-5; *See also* philosophy of physics
RealTime Physics, 21-8, 21-9; *See also* microcomputer-based laboratory
Reflective Judgment Interview, 11-5
Reflective Judgment Model, 11-4, 11-21
  epistemic cognition for, 11-15
  major periods and sample responses of, 11-6
Reflective Reasoning, 11-6
Reflective Writing (RW), 12-24–12-26
Reformed Teaching Observation Protocol (RTOP), 25-15
regression, 24-18; *See also* quantitative methods
  effects of nested data, 24-21
  hierarchical linear modeling, 24-21–24-22
  linear regression, 24-18
  logistic regression, 24-19
  maximum likelihood techniques, 24-20–24-21
relativism, 9-21–9-22, 11-7–11-8
replication, 9-25, 23-6; *See also* knowledge accumulation
  direct and conceptual, 23-8–23-9
  estimates of universal gravitation constant, 23-8
  estimation error, 23-9
  limits and exploiting variation among studies, 23-9–23-10
  programs of, 23-7–23-8
  reproducibility and, 23-7
  resources to facilitate, 23-10–23-16
  uncertainty in results, 23-9
reporting practices, 23-10; *See also* knowledge accumulation
  CONSORT checklist of impact study, 23-11–23-12
  description of intervention and comparison conditions, 23-15
  effect size reporting, 23-14
  general resources, 23-10–23-14
  publication bias, 23-15
  transparent and comprehensive reporting, 23-15–23-16

Representational Competence in Kinematics (KiRC), 20-14
Representational Fluency Survey (RFS), 20-14
Representational Variant of the Force Concept Inventory (R-FCI), 20-14, 20-21
representation of content, 16-22; *See also* textbook content analysis
  basic information, 16-22–16-23
  geographical distribution of affiliation of first author, 16-22
  multi-, 16-24
  readability of contents, 16-23–16-24
representations, 10-22
research-based (*RB*), 14-12
research-based conceptual instruments (RBIs), 24-2–24-4; *See also* quantitative methods
Research-Based Instructional Strategies (RBIS), 26-23
research-based teaching-learning sequences, 26-1
  design and evaluation of physics teaching learning sequences, 26-4–26-9
  design-based research, 26-2, 26-3, 26-9
  design-based research and evaluation of TLS, 26-22–26-23
  design process, 26-10–26-11
  design tools for intermediate frameworks, 26-13–26-21
  evaluation and redesign, 26-11–26-12
  exchange of educational practices and products, 26-21–26-22
  implementation, 26-11
  improvement of teaching materials, 26-3
  main ideas, 26-21
  retrospective analysis, 26-11–26-12
  teaching experiments, 26-11
  teaching learning sequences, 26-2
  TLS efficacy, 26-12
research design, 5-1, 5-24; *See also* equity in research design
  changes to, 5-23–5-24
researcher and researcher team, 5-4; *See also* equity in research design
  impact on research process, 5-5–5-6
  inviting additional researchers, 5-6–5-7
  safety and care for researchers and participants, 5-18–5-20
research experience for undergraduates (REU), 1-30
Research in Disabilities Education (RDE), 1-7
research instruments, 5-13; *See also* equity in research design
  considerations in designing, 5-13
  demographic questions on, 5-13–5-14
  in PER, 5-14–5-15
  understanding equity issues, 5-14–5-15
research participants, 5-10; *See also* equity in research design
  challenges in recruiting, 5-10–5-11
  as decision making partners in studies, 5-11–5-12
  safety and care for, 5-18–5-20
Research Programs, 9-10–9-11
research questions, 25-3; *See also* qualitative methods
  collecting data, 25-5–25-6
Romantic science, 12-9; *See also* philosophy of physics
root mean square error of approximation (RMSEA), 24-20

Schommer, M., 11-8–11-10
school curriculum, 15-1; *See also* curriculum
science, 9-13, 9-23, 10-23
  and knowledge, 9-4
  literacy, 15-19, 15-20
  Science and Engineering Practices, 6-6
  studies, 9-12–9-15
  textbooks, 16-2, 17-1
science education (SE), 9-3

Science Studies (SS), 9-2, 9-12, 9-23; *See also* physics, aims and values of
  causality principle, 9-13
  historians of science, 9-14
  human intentionality in scientific work, 9-14
  impartiality principle, 9-13
  knowledge, 9-13
  Knowledge and Social Imagery, 9-13
  natural philosophy, 9-14
  objectivity, 9-14, 9-15
  programmatic principles, 9-13
  scholars, 9-13
  science, 9-13
  scientism, 9-13
  teaching NOS away from CV, 9-24
Science, Technology, and Society (STS), 8-7, 16-17
Science, Technology, Engineering, Art and Mathematics (STEAM), 17-7
Science, Technology, Engineering, Mathematics, and Medicine (STEMM), 5-7
Science, Technology, Society, and Environment (STSE), 8-7
scientific; *See also* meta-curricular frameworks
  concepts in society, 7-6
  controversies, 12-5
  education goals, 26-1
  explanations, 15-23
  knowledge, 10-3
  literacy, 10-5
scientific inquiry (SI), 10-3, 10-5, 17-16; *See also* methods and practices in physics
  alternative approaches to NOS and, 10-7–10-8
  consensus-based approach to NOS and, 10-6–10-7
  domain-general and consensus-based view, 10-6
scientific literacy (SL), 7-1, 7-19–7-20, 10-5, 15-19, 15-20, 16-15–16-17; *See also* literacy
  acceleration, 7-16
  advances in physics of music and culture, 7-14
  aesthetic appreciation, 7-18
  aims and values, 7-15
  annual scientific production, 7-8
  annual trending topics for last two decades, 7-10
  artificial intelligence, 7-3
  astronomical revolution, 7-16
  beauty, 7-15
  communication with physicists and the public, 7-13
  concepts in AI and data science, 7-12
  depiction of object, 7-13
  depiction of retrograde motion using epicycles, 7-17
  digital tools, 7-2–7-3
  digital transformation, 7-5
  displacement, 7-16
  electrical and magnetic experiments, 7-13
  electric guitar, 7-14
  emerging new literacies in societies, 7-2–7-5
  emotions evoked by natural phenomena, 7-17
  empirical characteristics of light, 7-13
  epistemic knowledge, 7-11
  geocentric theory, 7-16
  historical direct citation network, 7-9
  intrinsic and extrinsic representation, 7-15
  as key in present and future, 7-5–7-7
  knowledge of science, 17-9
  physics as a key content of knowledge, 7-11–7-14
  physics as fundamental belief in, 7-14–7-19
  PISA science framework, 17-10, 17-11, 7-12
  procedural knowledge, 7-11
  quantitative analysis of SL content, 16-15–16-16
  quantum mechanics, 7-18
  radio broadcasting, 7-14
  scientific concepts in society, 7-6
  significance of, 7-1
  socio-scientific issues, 7-2
  sublimity, 7-17–7-18
  Sunday schools, 7-13–7-14

    symmetry, 7-16, 7-18
    themes used in senior secondary-school physics textbooks, 17-9
    trends in research on, 7-7–7-11
    vacuum valve, 7-14
    World Collaboration map about research on, 7-9
scientific method, 10-1–10-2; *See also* methods and practices in physics
    boundary work, 10-5
    causal demonstrations, 10-3
    as rhetorical tool, 10-5
    scientific and non-, 10-4
    scientific knowledge, 10-3
    scientific theories, 10-3
scientific theories, 10-3, 10-19; *See also* mathematization of physics; theoretical practices and methods
    modeling, 19-10
    objectivity, 18-12–18-13
    results and fundamental right, 9-2
    revolution, 12-6
scientism, 9-13
scree plot, 24-15
    for cluster analysis, 24-26
    of first 15 items on FCI, 24-16
second-order barrier, 1-23
Self-Determination Theory, 25-26; *See also* theoretical perspectives and frameworks
self-efficacy, 2-11
Semantic View of theories, 10-22, 10-23
sexism, microaggressions of, 4-6
sexist deviation, 8-15
Simple Knowledge, 11-9
simulations, 21-13; *See also* visualization and mathematization
    PhET simulations, 21-13–21-14
skill-theory, 11-3; *See also* epistemic beliefs
    hierarchies of skills, 11-3–11-4
Social Annotation Platform (SAP), 15-29
social issues and 21st century skills, 15-24–15-25
social justice (SJ), 8-2
social model of disability, 1-1
social network analysis (SNA), 24-27; *See also* network analysis
socio scientific issues (SSI), 7-2, 8-1, 8-5, 11-2; *See also* scientific literacy
    addressing SSI using HP, 8-15–8-16
    approach, 8-7
    cis-hetero concept of compulsory motherhood, 8-18
    Colonialist and Eurocentric values, 8-3
    considerations, 8-26–8-28
    conspiracy theory movements, 8-2
    Covid-19 pandemic, 8-2
    denial of prejudice, 8-17
    differences in educational attainments, 8-4
    division of positions, 8-20–8-26
    Enlightenment rationality, 8-13
    ethical frameworks, 8-10–8-11
    ethnic-racial disparities, 8-2
    Ethnomathematics program, 8-12–8-13
    experiential ethics of care, 8-13–8-14
    feminist criticism of science, 8-14–8-15
    gender identity, 8-17
    gender inequalities, 8-3
    and HP to promote social justice, 8-11–8-16
    inclusive attitude towards different points of view, 8-6–8-7
    Meitner and SSI on physics teaching, 8-16–8-26
    modes of thought, 8-9–8-10
    role of physics in establishing rationality, 8-13
    in science education, 8-5–8-10
    sexist deviation, 8-15
    socio-scientific controversies, 8-8
    strengthening of, 8-4–8-5
    STS theoretical framework, 8-7
    Thomas Hobbes and Robert Boyle, 8-6
    two elements in standpoint feminism, 8-19

Special Theory of Relativity (STR), 15-16
split-half method, 24-11
staging a teaching sequence tools, 26-17–26-21; *See also* design tools
Staging tools, 26-18; *See also* design tools
standard error of measurement (SEM), 24-11
standardized mean difference (SMD), 23-14
standardized test scores, 24-6
Standard Protocol Items: Recommendations for Intervention Trials (SPIRIT), 23-10
Standard Reference Data Act, 23-7
Statement, 10-22
stationary distribution. *See* Boltzmann distribution
statistical approaches to synthesizing effects, 23-5–23-6; *See also* knowledge accumulation
statistical methods, 24-4–24-9; *See also* quantitative methods
statistical power, 24-8–24-9; *See also* quantitative methods
STEM (science, technology, engineering, and mathematics), 1-2, 4-8, 8-3
  disabled students' experiences in postsecondary, 1-20–1-22
  goals for graduate education, 27-6
  LGBT research in, 3-8–3-13
  literature on STEM doctoral student retention, 4-4
  literature reviews of disability in, 1-6–1-9
  marginalized groups in, 4-1
  representation of disabled people in, 1-2–1-4
  research on cultural and structural context, 4-3
  socialization in, 2-17
  trends toward disabled people, 1-5–1-6
  21st century students, 27-6
  underrepresentation of disabled people in, 1-6
stereotype; *See also* equity and inclusion; gap-gazing
  of men and women in physics, 3-3
  threat, 2-5, 2-15–2-16
Strategies for Engaged Learning Framework (SELF), 2-4
structural equation modeling (SEM), 2-22, 24-22; *See also* equity and inclusion; quantitative methods
  CFI, 24-24
  methodology, 24-24
  models, 24-23
  path model, 24-22–24-24
  in PER studies, 24-25
Structure of Scientific Revolutions, The, 9-9, 9-10
Structure of the Observed Learning Outcomes (SOLOs), 11-3
student; *See also* design tools; equity in physics graduate education; philosophy of physics
  conceptions of difficulty, 26-17
  epistemology, 12-19–12-22
  experience in graduate school, 4-5–4-8
  retention, 4-4–4-5
  sense of belonging in physics courses, 2-12–2-13
Student Energy Targets (SETs), 19-18
sublimity, 7-17–7-18
Sunday schools, 7-13–7-14
super-symmetry (SUSY), 7-18
surveys, 25-12–25-13; *See also* qualitative methods
Sustainable Development Goals (SDG), 8-3, 16-20, 17-26
symbolic forms, 18-19–18-20, 21-5; *See also* mathematics in physics
  and notations, 18-13–18-14
symmetry, 7-16
Syntactic View of theories, 10-22, 10-23
system, 19-7; *See also* physics equations

teachers and teaching, 18-8–18-10; *See also* mathematics in physics
teaching assistants (TAs), 2-24
teaching interviews, 25-12; *See also* qualitative methods

Teaching Learning Sequences (TLS), 26-2; *See also* research-based teaching-learning sequences
  activity guide program, 26-6
  CLIS project, 26-5
  "communicative focus" tool, 26-6
  construction of TLS activities, 26-20
  DBR methodology, 26-3
  Developmental Research, 26-6
  Educational Reconstruction, 26-7
  framework into design brief, 26-18
  gaps in area of design-based research, 26-8–26-9
  grain size, 26-7–26-8
  ISLE approach, 26-7
  "knowledge in pieces" framework, 26-4
  Learning Demand tool, 26-6
  perspectives of design-based research and evaluation of, 26-22–26-23
  Piaget's genetic epistemology, 26-4
  research on design and evaluation of physics, 26-4–26-9
  theories and proposal product in revised approaches, 26-8
  "Tutorials", 26-5
  "Two Worlds", 26-6
technical pre-professional training (TPT), 12-2, 12-17
technology-enhanced PCK (TPCK), 19-18
temporary form, 26-16
Test of Calculus and Vectors in Mathematics and Physics (TCV-MP), 20-14
test of graphing in science (TOGS), 20-15
Test of Understanding Graphs in Calculus (TUG-C), 20-13–20-14
Test of Understanding Graphs in Kinematics (TUG-K), 20-12–20-13, 20-21
textbook, 1-15, 12-2, 12-3, 14-1, 15-2
  actions related to, 14-27
  analysis methods, 14-6
  authors/publisher expectations, 14-15
  categorization of research questions and gaps, 14-9
  chronological list of papers to be analyzed, 14-7
  content, 14-11, 14-28
  countries of PTER articles, 14-9–14-10
  creation recommendations, 14-22–14-23
  curriculum-maker expectations, 14-15
  discussion, 14-24
  education levels, 14-3–14-4
  expectation analysis methods, 14-15
  expectation holders, 14-3
  in extended conception of curriculum, 15-3–15-5
  form, 14-11, 14-29
  functions of physics textbook, 14-1
  historical phases, 14-4
  importance of textbooks, 15-4
  inductive categorization of gathered information, 14-9
  inductively gained categories of expectations, 14-10–14-12
  literature review limitations, 14-32
  methods, 14-5–14-9, 14-14–14-15, 14-27
  mistakes, 17-10–17-12
  nation-specific features, 14-4
  number of papers per country, 14-10
  outlook, 14-32–14-33
  pedagogy and instruction, 12-13–12-16
  in physics education, 14-1
  physics researcher expectations, 14-19–14-22
  recommendations, 14-22, 14-23–14-24, 14-31–14-32
  relationships between textbook research, creation, and use, 14-33
  research gaps, 14-12–14-13, 14-25
  research question, 14-5, 14-25–14-27
  research questions posed in analyzed papers, 14-13–14-14
  research recommendations, 14-24

textbook (*Continued*)
  results reported in analyzed papers, 14-15
  results, 14-9–14-24, 14-27–14-30
  search method, 14-5–14-6
  selection method, 14-6
  spreadsheet construction, 14-6–14-8
  spreadsheet conventions, 14-8
  structural elements of textbooks, 14-2–14-3
  student expectations, 14-18–14-19
  teacher expectations, 14-16–14-17
  textbook content, 15-3
  textbook evaluations, 15-2
  theory, 14-2
textbook and curriculum alignment, 15-1, 15-2, 15-30–15-32
  content, 15-3
  curriculum-based research, 15-9–15-18
  curriculum making and enacting, 15-1–15-2
  Enacted Curriculum, 15-5
  evaluations, 15-2
  four-level view of curriculum, 15-3
  importance of textbooks, 15-4
  initial contextual consideration, 15-5–15-9
  with intra-curricular processes, 15-26–15-30
  with meta-curricular frameworks, 15-19–15-26
  physics textbooks, 15-5–15-9
  place and role of, 15-3–15-5
  school curriculum, 15-1
  types of curricula, 15-2
textbook content analysis, 16-1, 16-27–16-29
  coding system, 16-4, 16-6
  conceptual level of content, 16-11–16-15
  content analysis of key techniques and methods, 16-24–16-27
  cultural level of content, 16-19–16-22
  education level of papers, 16-7–16-8
  epistemic level of content, 16-15–16-19
  findings on, 16-11–16-24
  geographical distributions of papers, 16-8–16-10
  limitations on sampling process of study, 16-29–16-31
  literature retrieval and screening, 16-3–16-4
  number of papers focused on physics textbooks, 16-2, 16-11
  process of paper selection for review, 16-5
  proportion of publications according to education level, 16-7
  publications according to education level, 16-8
  representation of content, 16-22–16-24
  research method, 16-3–16-6
  reviewed literature, 16-6
  search results in different databases, 16-4
textbooks, evaluation of physics, 17-1, 17-24–17-27; *See also* Physics Textbook Evaluation Literature
  development of PTEL, 17-5
  distribution over time of PTEL publication, 17-5
  domains, 17-3
  findings, 17-5
  geographical and historical aspects, 17-4
  key ideas, 17-2
  level of education, 17-4–17-5
  methods, 17-3–17-5
  online education resources, 17-2–17-3
  Physics Textbook Evaluation Literature, 17-3
  search strategy, 17-3–17-4
textbooks expectations research (TER), 14-26
theoretical entities, 9-20
theoretical perspectives and frameworks, 25-24; *See also* qualitative methods
  approaches to problem of theoretical framing, 25-25
  conceptual frameworks, 25-26
  ethnography, 25-28–25-29
  innovation decision-making process, 25-25
  issues addressed by, 25-26–25-29
  phenomenography, 25-28

practical frameworks, 25-25–25-26
Self-Determination Theory, 25-26
theoretical practices and methods, 10-19; *See also* methods and practices in physics
discovery myths, 10-19–10-20
Einstein, 10-19–10-20
Matthew–Matilda effect, 10-20–10-21
model, 10-21
Model Based View, 10-22
models and modeling, 10-21–10-23
neglect of Islamic influences on history of physics, 10-20
Newton, 10-19–10-20
representations and models, 10-22
scientific theories, 10-19
Semantic and Syntactic View of theories, 10-22, 10-23
theories
and hypotheses, 12-21–12-22
and instructional strategies, 12-19–12-26
to knowledge, 10-17
in qualitative research, 25-2
Semantic and Syntactic View of, 10-22, 10-23
Young and Fresnel wave theories, 12-23
Theory–Theory (TT), 12-19
The Physics Teacher (TPT), 15-7
thought
modes of, 8-9–8-10
simulations, 15-23–15-24
style, 9-11
thought experiments (TE), 15-23–15-24; *See also* meta-curricular frameworks
3-parameter logistic (3PL) model, 24-13
Three-State Law, 9-5
TIMSS (Trends in Mathematics and Science Study), 15-3
Tinto model, 5-20
Titius–Bode law, 12-21
tracker, 21-10, 21-11; *See also* video-analysis tools
Traweek's ethnographic work, 6-19
Tucker–Lewis Index (TLI), 24-20
"Tutorials", 26-5; *See also* Teaching Learning Sequences
21st century skills, 15-25–15-26
2-parameter logistic model (2PL), 24-12
2PL and the nominal response model (2PLNRM), 24-14
two-tier multiple-choice (2TMC), 15-14
"Two Worlds", 26-6; *See also* Teaching Learning Sequences
Tycho Brahe's model, 13-19
type I error, 24-8; *See also* quantitative methods
type II error, 24-8–24-9; *See also* quantitative methods

Ubuntu, 13-25–13-26
uncertainty in study results, 23-9
underrepresented demographic student groups, 2-1; *See also* equity and inclusion
physics education equity research, 2-3
unidimensional IRT, 24-13
United Nations Educational, Scientific and Cultural Organization (UNESCO), 8-3
United States (U.S.), 4-1
universal design for learning (UDL), 1-2, 1-27
-aligned strategies, 1-19–1-20
universe, clockwork, 12-4
unsettled state, 22-11
user-created simulations, 21-14–21-15; *See also* visualization and mathematization
GeoGebra, 21-15–21-16
U.S. Institute of Education Sciences (IES), 1-8, 23-3

vacuum valve, 7-14
values affirmation, 2-18–2-19
vectors, 18-16; *See also* mathematics in physics
Vergnaud's theory of conceptual fields, 18-23; *See also* mathematics in physics

Victorian Certificate of Education (VCE), 17-12, 17-9
video-analysis tools, 21-9; *See also* visualization and mathematization
  interactive software, 21-9
  tracker, 21-10, 21-11
  video source, 21-9
Vienna Circle, 9-6
Views About Science Survey (VASS), 11-10, 11-20, 11-23
Views About Scientific Inquiry (VASI), 10-7
Views of Nature of Science Questionnaire (V-NOS questionnaire), 9-19, 10-6, 10-8
Views of Scientific Inquiry (VOSI), 10-7
Views on Science–Technology–Society (VOSTS), 6-8
virtual reality (VR), 17-19, 21-5, 21-10
virtual simulations, 1-14
visual depictions in physics, 21-1
visualization, 17-16–17-19; *See also* Physics Textbook Evaluation Literature
visualization and mathematization, 21-1, 21-21
  augmented and mixed reality, 21-10–21-12
  bridging models and representations, 21-13–21-20
  bridging physical phenomena and formal representations, 21-8–21-12
  degrees of mathematization, 21-6
  educational games, 21-18–21-20
  external visualization, 21-2, 21-4
  free-body diagrams, 21-1
  history in physics education, 21-20–21-21
  history of visualization tools, 21-20–21-21
  internal visualization, 21-2
  kinematic graphs, 21-1
  mathematization cycle of physics, 21-7
  microcomputer-based laboratory tools, 21-8–21-9
  microworlds, 21-16–21-18
  simulations, 21-13–21-14
  symbolic forms, 21-5
  theories of, 21-3–21-7
  theories of mathematization, 21-5–21-6
  theories of visualization, 21-4–21-5
  user-created simulations, 21-14–21-15
  video-analysis tools, 21-9–21-10, 21-11
  visualization and mathematization, 21-6–21-7
  visualization tools, 21-3, 21-7
  visualization tools in PER, 21-20

Warwick's study of training in mathematics, 6-18
wave-particle duality, 12-23
Web Content Accessibility Guidelines (WCAG), 1-4
What Works Clearinghouse (WWC), 23-3
Whig-history, 10-9; *See also* history of physics
Women of Color (WOC), 13-21
work–life balance and mental health, 4-8; *See also* equity in physics graduate education
Workshop Physics, 21-8; *See also* microcomputer-based laboratory
World Health Organization (WHO), 1-1
World Report on Disability, 1-3
World Wide Web (WWW), 17-22
World Wide Web Consortium (W3C), 1-4

Young and Fresnel wave theories, 12-23

Zeeman effect, 7-12
zone of proximal development (ZPD), 26-16

The International Handbook of Physics Education Research: Special Topics covers the topics of equity and inclusion; history and philosophy of physics; textbooks; mathematics; research history, methodologies, and themes. As the field of physics education research grows, it is increasingly difficult for newcomers to gain an appreciation of the major findings across all sub-domains, discern global themes, and recognize gaps in the literature. *The International Handbook of Physics Education Research: Special Topics* incorporates the understanding of both physics and education concepts and provides an extensive review of the literature in a wide range of important topics.

The International Handbook of Physics Education Research: Special Topics includes:

- The history and philosophy of physics teaching, including a review of physics textbooks.
- Teaching mathematics for physics students.
- Methodologies in physics education research and the future of physics departments.

Readers will find this comprehensive treatment of the literature useful in understanding physics education research and extending to all the physical sciences including chemistry, mathematics, astronomy, and other related disciplines.

Mehmet Fatih Taşar, Ph.D., of Georgia State University is a physics education researcher who has contributed to the field in various lead capacities. He taught physics and conducted research at many prestigious institutions including Penn State University, RMIT University, Gazi University, and Georgia State University. He has supervised graduate students to successful completion of their degrees and organized many international conferences such as *WCPE*, *ESERA*, and *iSER*. Dr. Taşar has served as an editor for scholarly journals including *EJMSTE*, *IJPCE*, *Hellenic Journal of STEM*, and *ARISE*.

Paula R. L. Heron, Ph.D., is Professor of Physics at the University of Washington where she has been teaching since 1995. She is co-founder and co-chair of the biannual “Foundations and Frontiers in Physics Education Research” conference series, and she has been a committee executive for groups within the American Physical Society (APS), American Association of Physics Teachers (AAPT), and National Research Council. She serves as an associate editor of *Physical Review* (APS). She was elected a Fellow of APS in 2007 and 2008.

AAPT PHYSICS EDUCATION®

AIP Publishing

9 780735 425484

Not for resale