Springer International Handbooks of Education

SPRINGER REFERENCE

Dani Ben-Zvi · Katie Makar
Joan Garfield *Editors*

International Handbook of Research in Statistics Education

EXTRAS ONLINE

Springer

Springer International Handbooks of Education

More information about this series at http://www.springer.com/series/6189

Dani Ben-Zvi • Katie Makar • Joan Garfield
Editors

International Handbook of Research in Statistics Education

Springer

Editors
Dani Ben-Zvi
Faculty of Education
The University of Haifa
Haifa, Israel

Katie Makar
School of Education
University of Queensland
St Lucia, QLD, Australia

Joan Garfield
Department of Educational Psychology
The University of Minnesota
Minneapolis, MN, USA

ISSN 2197-1951 ISSN 2197-196X (electronic)
Springer International Handbooks of Education
ISBN 978-3-319-66193-3 ISBN 978-3-319-66195-7 (eBook)
https://doi.org/10.1007/978-3-319-66195-7

Library of Congress Control Number: 2017959604

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

This handbook is a collection of articles, grounded in science but equally dedicated to practice, on topics carefully chosen by the editors and written by leading authors in the field. At the same time, this book is much more than just a collection of articles. It is the realization of a decades-old vision for the future of a new subject, the flowering of seeds planted long ago and cultivated by generations of students and scholars.

In the beginning, there was the soil and the seed. The soil was statistics, the science of learning from data, not a crop in itself, so much as an environment for raising a crop, using data to learn about a subject area. The seeds came from the subject area, in this case education, wanting to learn how we learn. When the seed found the right soil, there sprouted the new science of statistics education, driven by a commitment to teaching statistics well and led in the early years by such statistics textbooks as Snedecor and Cochran (1937), Hoel (1947), Hogg and Craig (1958), and Mosteller, Rourke, and Thomas (1961) and, more recently, in a second wave, by Freedman, Pisani, and Purvis (1978), Moore (1978), Anscombe (1981), Moore and McCabe (1989), and Scheaffer, Watkins, Witmer, and Gnanadesikan (1996). In parallel with the second wave, a new academic subject was born: using statistics to learn how to teach statistics.

Like farming, teaching is a craft, not an innate talent you either have or don't have, not a green thumb you are born with or without, but something you can learn from experience if you pay attention to data. Data from research in statistics education can and should guide how those of us who teach statistics can improve our craft of teaching statistics well. At the same time, we who teach statistics can and should guide those who use data for their study of how students learn. In that spirit, the authors who have written here have much to offer to those of us who teach the science of learning from data to use data to improve how we teach. Surely, statistics education stands at a junction of uniquely fruitful possibilities.

The advance of *statistics as a subject* depends on the advance of *statistics as a profession*. The advance of statistics as a profession depends on the advance of *teaching statistics as a calling*. The advance of statistics teaching as a calling depends on *teaching statistics as a craft*, and the advance of teaching statistics as a craft depends on *scientific research on how students learn* statistics and how teachers who pay attention can best help their students learn.

I conclude my preface with a concerning challenge and a brief for optimism. I start with my concerning challenge: a growing time lag. The cutting edge of statistics-the-subject is advanced by those who do the new research and teach it to their graduate students in their Ph.D. programs. But that reach is narrow—the person who teaches the newest ideas in genomics is not often the same person who teaches the latest in Markov chain Monte Carlo or mining of business data. The cutting edge of teaching statistics-the-subject is advanced mainly by the very few Ph.D. graduates from those research-oriented universities who choose to teach at the few liberal arts colleges where the lighter teaching load allows time for curricular innovation and carries an expectation of creating new courses. A cutting edge of research in statistics education, one edge among many, depends on those faculty from liberal arts colleges who write and give talks about curricular innovations that trickle down from the research universities. To exploit the pernicious metaphor of trickle down: innovations in statistics-the-subject are funded from the top, innovators in the K-16 curriculum get the fewer and smaller grants in statistics education, and the researchers in statistics education have all too often been consigned to feed off what's left at the bottom. That was then. Fortunately, the growing emphasis on data science and assessment of effectiveness is changing these priorities.

My point here is to illustrate a concern about a time lag. It takes time for the latest research to make its way into graduate courses, and thence to the undergraduate curriculum, and from there to articles and presentations that engage the attention of those education researchers who use data science to advance our understanding of how students learn data science. It's not just the current time lag. I worry even more about a possible growing divergence between statistics-the-subject and statistics education research. Until recently, the overlap between statistics-the-subject and what education researchers learned about statistics as part of their Ph.D. programs could be regarded as the core of the subject. Now, statistics-the-subject expands rapidly in many directions. My concern: what should statistics education research regard as the core?

Concern and challenge aside, I conclude with deep reasons for optimism. Research in statistics education is unique in that the target subject (statistics), whose teaching and learning is being studied, is at the same time the main conceptual and methodological approach to learning from the research data. Those who study the teaching and learning of chemistry do not rely primarily on molecules to understand how their students learn. Those who study the learning and teaching of astronomy or microbiology do not observe students through a telescope or microscope. Uniquely in statistics education, all three of the (1) subject itself, (2) those committed to teaching the subject well, and (3) those who use science to study the teaching and learning of the subject share a "common core," the subject itself. Our history bears this out: there have been no "stat wars."

It is a deep pleasure to recommend this pioneering and peaceful volume to your attention.

George W. Cobb
Mount Holyoke College ,
South Hadley, MA, USA

References

Anscombe, F. J. (1981). *Computing in statistical science through APL*. New York: Springer.

Freedman, D., Pisani, R., & Purves, R. (1978). *Statistics* (1st Ed.). New York, NY: W.W. Norton & Company.

Hoel, P. G. (1947). *Introduction to mathematical statistics* (1st Ed.). Wiley.

Hogg, R. V., & Craig, A. T. (1958). *Introduction to mathematical statistics* (1st Ed.). New York: Macmillan Publishing.

Moore, D. S. (1978). *Statistics: Concepts and controversies* (1st Ed.). New York: W. H. Freeman and Company.

Moore, D. S., & McCabe, G. P. (1989). *Introduction to the practice of statistics* (1st Ed.). New York: W. H. Freeman and Company.

Mosteller, F., Rourke, R. E. K., & Thomas, G. B. (1961). *Probability with statistical applications* (1st Ed.). Reading, MA: Addison-Wesley.

Scheaffer, R. L., Watkins, A., Witmer, J., & Gnanadesikan, M. (1996). *Activity-based statistics: Instructor's guide* (1st Ed.). New York: Springer.

Snedecor, G. W., & Cochran, W. G. (1937). *Statistical methods* (1st Ed.). Iowa University Press.

Preface

After many years of hard work and extensive collaboration, we are deeply honored to present to the statistics and mathematics education communities the first *International Handbook of Research in Statistics Education*. We have been working for several decades to bring visibility and credibility to this important discipline that supports the teaching and learning of statistics. This handbook not only reflects those efforts but is designed to further promote high-quality research and improvements in the teaching and learning of statistics.

This book builds on our commitment over the past decade to explore ways to understand and develop students' statistical literacy, reasoning, and thinking. Despite living and working in three different countries, we have collaborated together in person and via frequent Skype calls to produce a volume that reflects the current knowledge and ideas in statistics education.

Initial conversations about the need for a handbook began at the 2011 gathering of the International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL) held in the Netherlands. Plans solidified and an initial editorial board meeting was held in conjunction with the 2013 SRTL meeting in the USA. We created a structure that included three main sections of the book, each overseen by two coeditors. Together we shaped the scope and goals of a unique handbook that could provide a valuable foundation for educators and researchers. We are deeply indebted to the six editors who worked with us in helping this vision become a published book:

Part I: Beth Chance and Allan Rossman
Part II: Maxine Pfannkuch and Robert delMas
Part III: Janet Ainley and Dave Pratt

It has been a great experience to work on this book with such dedicated and top-notch editors and with the group of international scholars who collaborated on each chapter. Producing this book required ongoing reading, writing, discussing, and learning as well as a face-to-face meeting with many of the authors at ICOTS in the USA (July 12–13, 2014).

We gratefully acknowledge the tremendously valuable assistance and feedback offered by all of the editors and chapter authors who reviewed the book chapters and from the entire SRTL community. We would also like to thank our colleagues in statistics education who acted as external reviewers for the chapters: Dor Abrahamson, Keren Aridor, Pip Arnold, Arthur Bakker, Stephanie Budgett, Gail Burrill, Helen Chick, Neville Davies, Adri Dierdorp, Andreas Eichler, Lyn English, Jill Fielding-Wells, Iddo Gal, Einat Gil, Randall Groth, Jennifer Kaplan, Sibel Kazak, Cliff Konold, Gillian Lancaster, Cindy Langrall, Aisling Leavy, Rich Lehrer, Marsha Lovett, Helen MacGillivray, Sandy Madden, Hana Manor Braham, Maria Meletiou-Mavrotheris, Jen Noll, Susan Peters, Robyn Pierce, Robyn Reaburn, Jackie Reid, Jim Ridgway, Luis Saldanha, Susanne Schnell, Mike Shaughnessy, Bob Stephenson, Jane Watson, Jeff Wilmer, Lucia Zapata-Cardona, and Andy Zieffler.

It has been a positive and productive collaboration. It has also been a delight for the three of us to work together on this project, especially as one of us (Joan Garfield) retires from her faculty position.

We are grateful to Springer Publishers, for providing a publishing venue for this book, and to Joseph Quatela, the editor who skillfully managed the publication on their behalf.

Lastly, many thanks go to our spouses Hava Ben-Zvi, Sanjay Makar, and Michael Luxenberg and to our children—Noa, Nir, Dagan, and Michal Ben-Zvi, Keya Makar, and Harlan and Rebecca Luxenberg—as our primary sources of energy and support.

Dani Ben-Zvi Haifa, Israel
Katie Makar St. Lucia, QLD, Australia
Joan Garfield Minneapolis, MN, USA

Contents

Contributors

Janet Ainley School of Education, University of Leicester, Leicester, UK

Pip Arnold Cognition Education, Auckland, New Zealand

James Baglin RMIT University, Melbourne, VIC, Australia

Arthur Bakker Freudenthal Institute, Utrecht University, Utrecht, The Netherlands

Dani Ben-Zvi Faculty of Education, The University of Haifa, Haifa, Israel

Rolf Biehler Institute of Mathematics, University of Paderborn, Paderborn, Germany

Beth Chance Department of Statistics, California Polytechnic State University, San Luis Obispo, CA, USA

George W. Cobb Mt. Holyoke College, Hadley, MA, USA

Jere Confrey College of Education, North Carolina State University, Raleigh, NC, USA

João Pedro da Ponte Instituto de Educação, Universidade de Lisboa, Lisboa, Portugal

Robert delMas Department of Educational Psychology, University of Minnesota, Minneapolis, MN, USA

Lyn English Faculty of Education, Queensland University of Technology, Brisbane, QLD, Australia

Jill Fielding-Wells School of Education, The University of Queensland, Brisbane, QLD, Australia

Noleine Fitzallen Faculty of Education, University of Tasmania, Hobart, TAS, Australia

Daniel Frischemeier Institute of Mathematics, University of Paderborn, Paderborn, Germany

Elizabeth Fry Department of Educational Psychology, University of Minnesota, Minneapolis, MN, USA

Iddo Gal Department of Human Services, The University of Haifa, Haifa, Israel

Joan Garfield Department of Educational Psychology, The University of Minnesota, Minneapolis, MN, USA

Robert Gould Department of Statistics, University of California, Los Angeles, CA, USA

Koeno Gravemeijer Eindhoven University of Technology, Eindhoven, Netherlands

Randall Groth Education Specialty Department, Salisbury University, Salisbury, MD, USA

Nicholas J. Horton Department of Mathematics and Statistics, Amherst College, Amherst, MA, USA

Ryan Seth Jones College of Education, Middle Tennessee State University, Murfreesboro, TN, USA

Sibel Kazak Faculty of Education, Pamukkale University, Denizli, Turkey

Frauke Kreuter University of Maryland, College Park, MD, USA

University of Mannheim, Mannheim, Germany

Hollylynne S. Lee College of Education, North Carolina State University, Raleigh, NC, USA

Richard Lehrer Department of Teaching and Learning, Peabody College, Vanderbilt University, Nashville, TN, USA

Sandra Madden Teacher Education and Curriculum Studies Department, University of Massachusetts Amherst, Amherst, MA, USA

Katie Makar School of Education, The University of Queensland, Brisbane, QLD, Australia

Kevin McConway School of Mathematics and Statistics, The Open University, Milton Keynes, UK

Amelia McNamara Smith College, Northampton, MA, USA

Maria Meletiou-Mavrotheris School of Humanities, Social and Education Sciences, European University Cyprus, Egkomi, Cyprus

Per Nilsson School of Science and Technology, Örebro University, Örebro, Sweden

Jennifer Noll Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR, USA

Roger D. Peng Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA

Peter Petocz Department of Statistics, Macquarie University, North Ryde, NSW, Australia

Maxine Pfannkuch Department of Statistics, The University of Auckland, Auckland, New Zealand

Dave Pratt Institute of Education, University College London, London, UK

Randall Pruim Department of Mathematics and Statistics, Calvin College, Grand Rapids, MI, USA

Chris Reading SiMERR National Research Centre, University of New England, Armidale, NSW, Australia

Anna Reid Sydney Conservatorium of Music, University of Sydney, Sydney, NSW, Australia

Jim Ridgway School of Education, Durham University, Durham, UK

Allan Rossman Department of Statistics, California Polytechnic State University, San Luis Obispo, CA, USA

Andee Rubin TERC, Cambridge, MA, USA

Maike Schindler Faculty of Human Sciences, University of Cologne, Cologne, Germany

J. Michael Shaughnessy The Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR, USA

Jessica M. Utts Department of Statistics, University of California—Irvine, Irvine, CA, USA

Jane Watson Faculty of Education, University of Tasmania, Hobart, TAS, Australia

Christopher J. Wild Department of Statistics, The University of Auckland, Auckland, New Zealand

Jeff Witmer College of Arts and Sciences, Oberlin College, Oberlin, OH, USA

Andrew Zieffler Department of Educational Psychology, The University of Minnesota, Minneapolis, MN, USA

Introduction

Statistics education has come of age. This unique discipline has emerged from a diverse set of foundations. Over the past 50 years, it has grown and developed its own identity. Every discipline needs a recognized body of research to establish its credibility as a legitimate field of knowledge and study. It is now time for this relatively new field of statistics education to have a research handbook that provides a collection and synthesis of the body of knowledge that supports the teaching and learning of statistics.

Statistics has become one of the most central topics of study in the modern world of information and big data. The dramatic increase in demand for learning statistics in all disciplines is accompanied by tremendous growth in research in statistics education. Increasingly, educators at all levels are teaching more topics and courses in quantitative reasoning, data analysis, and data science at lower and lower grade levels within mathematics and science and across other content areas. However, despite the growth in statistics education, research has continually revealed many challenges in helping learners develop statistical literacy, reasoning, and thinking. New curricula and technology tools show promise in facilitating the achievement of these desired outcomes. New research in the field can critically inform college instructors, classroom teachers, curriculum designers, researchers in mathematics and statistics education, policymakers, and newcomers to the field of statistics education.

This *International Handbook of Research in Statistics Education* aims to provide a solid foundation for such studies. Like statistics itself, statistics education research is by nature interdisciplinary, with its practices and principles developed from many different fields. In addition, current problems and methods of statistical practice in the changing world need to be shared with educators who are teaching today's students.

We see the foundations of the knowledge and research findings presented in this handbook as based primarily on three areas of work that can be represented by the contributions of three extraordinary individuals who have made major contributions to knowledge. They are John Tukey (1915–2000), Amos Tversky (1937–1996), and Jean Piaget (1886–1980).

Tukey helped move the practice of statistics into a new era of exploring data. In the 1970s, the reinterpretation of statistics into separate practices comprising exploratory data analysis (EDA) and confirmatory data analysis (CDA, inferential statistics) (Tukey, 1977) freed certain kinds of data analysis from ties to probability-based models, so that the analysis of data could begin to acquire status as an independent intellectual activity. The introduction of simple data tools, such as stem and leaf plots and boxplots, paved the way for students at all levels to analyze real data interactively without having to spend hours on the underlying theory, calculations, and complicated procedures. The work of Tukey and his colleagues was spread to students at all levels and led to new curriculum at the primary and secondary school. Computers and new pedagogies would later complete the "data revolution" in statistics education.

Tversky studied and enlightened the world about the ways people think and reason. He documented how often people misunderstand randomness and probability, leading them to use faulty heuristics when reasoning about samples. The work of Tversky and his colleagues (e.g., Kahneman, Slovic, & Tversky, 1982) has led to recognition of the new ways to build learning trajectories on existing foundations, of challenging faulty heuristics and biases through simulations and visual explorations, and of carefully assessing reasoning and thinking.

Finally, the work of Piaget (e.g., Inhelder & Piaget, 1958; Piaget & Inhelder, 1962) provided models of how to carefully study individual children as they understand the world, as well as how they reason about chance and probability (Piaget & Inhelder, 1975). His methods for studying students in depth have paved the way for current researchers who carefully observe and study the thinking of children, including how they think about inference and chance events.

Although their work was developed in the last century, we believe that much of the current research in statistics education has been built on the insights and knowledge that these three brilliant and creative thinkers provided—and statistics education may be unique in connecting the fruits of their studies. Research in statistics education includes studies of how people think about data and chance, the errors they systematically make that affect their inferences and judgments, the use and impact of new tools and learning environments, and the use of rich, qualitative data through observation interviews and teaching experiments.

The work in this volume represents a collaboration amid a diverse set of professionals, including leading educators, researchers, and statisticians from around the world. Our goal was to provide a resource that connects the practice of statistics to the teaching and learning of the subject in light of current and future challenges. The chapter authors and part editors contributed in the development, writing, and editing of this book. Just as the discipline of statistics education is built on diverse scientific areas, the writers of these chapters come from the departments of statistics, mathematics education, psychology, and technology education and, in some cases, new programs in statistics education.

We have organized the book into three main parts that encompass the breadth and depth of research on teaching and learning statistics across educational levels and settings. As such, the authors and editors strove to work in collaboration to link the main sections of the book with the diversity of ideas articulated throughout this handbook.

Part I: *Statistics, Statistics Education, and Statistics Education Research*

Part I of this handbook describes the interplay among the disciplines of statistics, statistics education, and statistics education research.

Part II: *Major Contributions of Statistics Education Research*

This part focuses on major contributions of statistics education research that relate to teaching, learning, and understanding statistics. It includes summaries of classic work as well as current work to show the progress and contrasting perspectives on main themes. Gaps in the research knowledge base are also identified.

Part III: *Contemporary Issues and Emerging Directions*

The focus of Part III of the handbook is on looking forward and examining emerging areas of statistics education research and their implications. Much of this section discusses more theoretical than empirical findings as the topics often have little research published so far or may be anticipated to be "on the horizon".

The collaboration that led to the production of this handbook aims to provide a resource that can be utilized by all people interested in the latest international research on teaching and learning statistics. We are proud to be part of this new discipline that has come of age, and we look forward to seeing new advances in the teaching and learning of statistics to students at all levels.

Dani Ben-Zvi
The University of Haifa,
Haifa, Israel

Katie Makar
The University of Queensland, St Lucia
QLD, Australia

Joan Garfield
The University of Minnesota, Minneapolis
MN, USA

References

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic Books.

Kahneman, D., Slovic, P., Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Piaget, J., & Inhelder, B. (1962). *The psychology of the child*. New York: Basic Books.

Piaget, J., & Inhelder, B. (1975). The origin of the idea of chance in children. UK: Routledge & Kegan Paul Ltd.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Part I
Statistics, Statistics Education, and Statistics Education Research

Beth Chance and Allan Rossman

Introduction to Part I

Our goal with Part I of the *International Handbook of Research in Statistics Education* is to set the stage for the articles and topics that form the bulk of the collection. We do this with three chapters that overview the history, core components, and future of the fields of statistics, statistics education, and statistics education research. In other words, Part I provides a careful examination of the interplay of the last three words of the handbook's title. Consistently enough, all three chapters are written by a team of three coauthors, experienced and prominent statisticians, statistics educators, and statistics education researchers representing four countries. The goal of Part I is to give the reader an understanding of how statistics education has developed and grown into a discipline of its own, with an eye to future needs and research questions.

We hope that this part of the handbook will help to establish common ground and encourage interaction among three groups that are integral to statistics education: statisticians, teachers of statistics, and statistics education researchers. These groups have much to offer each other. For example, teachers can improve their students' learning of statistics not only by understanding the discipline but also by applying the findings of education researchers and adapting them for effective teaching. Similarly, education researchers can engage in more meaningful and impactful research studies by not only knowing the discipline but also noting practices of effective teachers. We fear that even when members of these groups attend the same conference, such as ICOTS (the International Conference on Teaching Statistics), they may tend to focus on their own sessions and not engage with members of other groups; we therefore hope that the three chapters in this part lead to increased communication and consultation. Data scientists comprise a fourth group to be invited to the conversation, and these chapters touch on the role of data science in statistics education and its future.

To help understand the nature of statistics education research, we believe it is important to first understand the nature of statistics itself and how it differs from other academic disciplines, which helps inform what should be taught and how. To address this very broad topic in the opening chapter, we turned to three accomplished statisticians who have also been very involved with curriculum development and other statistics education efforts at tertiary and secondary levels. Chris Wild has served as president of the International Association for Statistics Education and has provided substantial input into the K-12 statistics curriculum for New Zealand. Jessica Utts has been president of the American Statistical Association and has written well-received textbooks for introductory statistics at the undergraduate level. Nick Horton chaired the ASA committee that revised guidelines for undergraduate programs in statistics and also served on the committee that revised the GAISE (Guidelines for Assessment and Instruction in Statistics Education) report for introductory undergraduate statistics. All three authors have played a role in identifying the needs and agenda for statistics education for informed consumers of quantitative information.

The first chapter, written by Wild, Utts, and Horton, discusses the nature of statistics as a scientific discipline, recounting some of its history and identifying characteristics that exemplify statistical thinking and draw distinctions with mathematical thinking. The chapter also considers why statistics is important and relevant in a wide variety of professional fields as well as for all educated citizens. Several examples highlight how statistical thinking is relevant in everyday life and necessary for advancing knowledge. The chapter concludes with discussion of statistics as a still-evolving and vibrant discipline in our contemporary world where data abound. Noting the growth of statistics as an academic major and as a career choice, along with the emergence of data science as a closely related field and popular career option, the final section dares to make some predictions about where the discipline of statistics is heading in the coming decades.

Having addressed the question of what is statistics, the second chapter moves on to consider the question: What is Statistics Education? The three authors share an association with the first doctoral program in the USA to specialize in statistics education. Joan Garfield is one of the leading statistics education researchers of the past 25 years, whose research and writings have had a substantial impact on the teaching and learning of statistics at all levels. Joan played a leading role in developing and shaping the Ph.D. program in Statistics Education at the University of Minnesota. Two of her former students in that program have joined her in authoring Chap. 2. Andy Zieffler was Joan's first student in this program and has gone on to lead curriculum development and other statistics education projects. Elizabeth Fry is one of Joan's most recent students, whose research focuses on developing high-quality assessments of curricular innovation in statistics education.

Chapter 2, written by Zieffler, Garfield, and Fry, begins by providing a brief history about the teaching and learning of statistics, first grounded in mathematics and science education and later broadened to include the teaching of statistics for all students at all levels. The authors identify milestones that led to statistics education establishing itself as a viable academic discipline, separate from statistics and from mathematics education. These milestones include the establishment of various professional associations, conferences, and journals devoted exclusively to statistics education. Statistics education encompasses a wide breadth of aspects and topics that will appear in later chapters in this volume. Zieffler et al. discuss several of these, including cognitive and noncognitive instructional goals, pedagogical approaches, use of technology, teacher preparation, and graduate programs in statistics education. As much as possible in a single chapter, these issues are discussed for various education levels and from an international perspective.

As statistics education has evolved and distinguished itself, so has the related research literature. The final chapter in Part I provides perspectives on different forms of that research, and how that research has been positively influenced from several converging disciplines. The author team includes two recent editors of the *Statistics Education Research Journal*: Peter Petocz and Iddo Gal, and a third expert in both qualitative and quantitative approaches to education research: Anna Reid.

As one piece of this chapter, Petocz, Reid, and Gal performed their own research study to examine the current landscape of statistics education research by conducting a qualitative study of articles published in various outlets for statistics education research in the past few years. Their goals are to describe who performs this kind of research, what kinds of questions are being addressed with this research, and the methods and conceptual schemes used to conduct the research. Petocz et al. distinguish between what they refer to as small-r and large-R research, the former addressing local problems in a particular context and the latter investigating larger issues that generalize more broadly. After analyzing the results of their analysis, they conclude the chapter by suggesting some directions for future development of the discipline of statistics education.

This first part of the handbook addresses very far-reaching themes as it strives to provide a context for the more focused chapters that comprise the other parts of the handbook. We expect these opening three chapters to provide the reader with valuable perspectives on the challenging questions of what constitutes statistics, statistics education, and statistics education research. We trust that these chapters will succeed in whetting the reader's appetite to discover the research findings in statistics education described in remaining handbook chapters, understand their implications, and look towards the future of statistics education.

Chapter 1
What Is Statistics?

Christopher J. Wild, Jessica M. Utts, and Nicholas J. Horton

Abstract What is statistics? We attempt to answer this question as it relates to grounding research in statistics education. We discuss the nature of statistics as the science of learning from data, its history and traditions, what characterizes statistical thinking and how it differs from mathematics, connections with computing and data science, why learning statistics is essential, and what is most important. Finally, we attempt to gaze into the future, drawing upon what is known about the fast-growing demand for statistical skills and the portents of where the discipline is heading, especially those arising from data science and the promises and problems of big data.

Keywords Discipline of statistics • Statistical thinking • Value of statistics • Statistical fundamentals • Decision-making • Trends in statistical practice • Data science • Computational thinking

1.1 Introduction

In this, the opening chapter of the *International Handbook on Research in Statistics Education*, we ask the question, "What is statistics?" This question is not considered in isolation, however, but in the context of grounding research in statistics education. Educational endeavor in statistics can be divided very broadly into "What?" and "How?" The "What?" deals with the nature and extent of the discipline to be conveyed, whereas any consideration of "How?" (including "When?") brings into

C.J. Wild (✉)
Department of Statistics, The University of Auckland, Auckland, New Zealand
e-mail: c.wild@auckland.ac.nz

J.M. Utts
Department of Statistics, University of California—Irvine, Irvine, CA, USA
e-mail: jutts@uci.edu

N.J. Horton
Department of Mathematics and Statistics, Amherst College, Amherst, MA, USA
e-mail: nhorton@amherst.edu

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_1

play a host of additional factors, such as cognition and learning theories, audience and readiness, attitudes, cultural issues, social interactions with teachers and other students, and learning, teaching and assessment strategies and systems.

This chapter discusses the nature of statistics as it relates to the teaching of statistics. The chapter has three main sections. Section 1.2 discusses the nature of statistics, including a brief historical overview and discussion of statistical thinking and differences with mathematics. Section 1.3 follows this up with a discussion of why learning statistics is important. Section 1.4 concludes the chapter with a discussion of the growing demand for statistical skills and a look at where the discipline is heading. A thread pervading the chapter is changing conceptions of the nature of statistics over time with an increasing emphasis recently on broad, as opposed to narrow, conceptions of what statistics is. We emphasize broader conceptions because we believe they best address emerging areas of need and because we do not want researchers to feel constrained when it comes to deciding what constitutes fair game as targets for research.

1.2 The Nature of Statistics

"Statistics"—as defined by the American Statistical Association (ASA)—"***is the science of learning from data, and of measuring, controlling and communicating uncertainty."*** Although not every statistician would agree with this description, it is an inclusive starting point with a solid pedigree. It encompasses and concisely encapsulates the "wider view" of Marquardt (1987) and Wild (1994), the "greater statistics" of Chambers (1993), the "wider field" of Bartholomew (1995), the broader vision advocated by Brown and Kass (2009), and the sets of definitions given in opening pages of Hahn and Doganaksoy (2012) and Fienberg (2014).

Figure 1.1 gives a model of the statistical-inquiry cycle from Wild and Pfannkuch (1999). This partial, rudimentary "map" hints at the diversity of domains that contribute to "learning from data." The ASA description of statistics given above covers all elements seen in this diagram and more. Although statisticians have wrestled with every aspect of this cycle, particular attention has been given by statistical theory-and-methods thinkers and researchers to different elements at different times. For at least the last half century, the main focus has been on the use of probabilistic models in the analysis and conclusion stages and, to a lesser extent, on sampling designs and experimental designs in the plan stage. But a wider view is needed to chart the way of statistics education into the future.

The disciplines of statistics and, more specifically, statistics education are, by their very nature, in the "future" business. The mission of statistics education is to provide conceptual frameworks (structured ways of thinking) and practical skills to better equip our students for their future lives in a fast-changing world. Because the data universe is expanding and changing so fast, educators need to focus more on looking forward than looking back. We must also look back, of course, but predominantly so that we can plunder our history's storehouses of wisdom to better chart

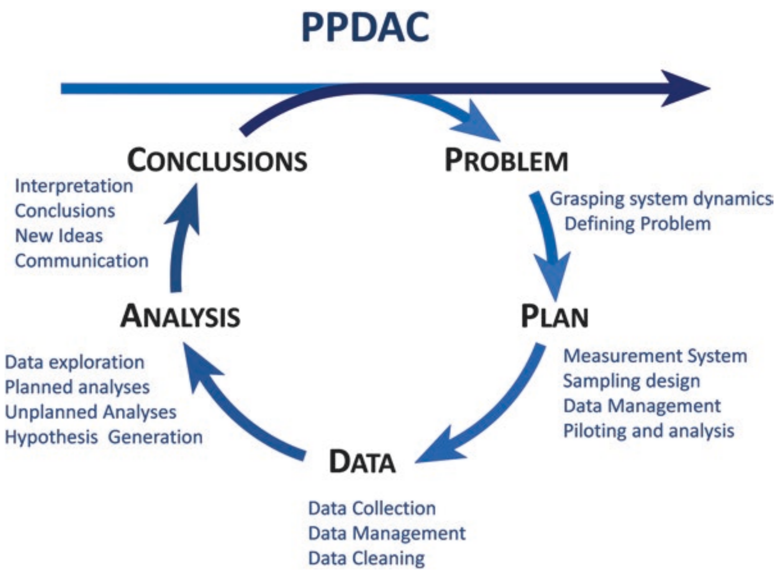


Fig. 1.1 The statistical-inquiry cycle

pathways into the future. For educational purposes, statistics needs to be defined by the ends it pursues rather than the means statisticians have most often used to pursue them in the past. Changing capabilities, like those provided by advancing technology, can change the preferred means for pursuing goals over time, but the fundamental goals themselves will remain the same. The big-picture definition that we opened with "keeps our eyes on the ball" by placing at the center of our universe the fundamental human need to be able to learn about how our world operates using data, all the while acknowledging sources and levels of uncertainty:

> Statisticians develop new methodologies in the context of a specific substantive problem, but they also step back and integrate what they have learned into a more general framework using statistical principles and thinking. Then, they can carry their ideas into new areas and apply variations in innovative ways. (Fienberg, 2014, p. 6)

At their core, most disciplines think and learn about some particular aspects of life and the world, be it the physical nature of the universe, living organisms, or how economies or societies function. Statistics is a meta-discipline in that it *thinks about how to think* about turning data into real-world insights. Statistics as a meta-discipline advances when the methodological lessons and principles from a particular piece of work are abstracted and incorporated into a theoretical scaffold that enables them to be used on many other problems in many other places.

1.2.1 History of Statistics

This section outlines the evolution of major threads that became interwoven to make statistics what it is today, forming the basis of the ways in which we gather, think about, and learn using data. These threads include the realization of the need for data, randomness and probability as a foundation for statistical modeling and dealing with uncertainty, theories underpinning principled approaches to data collection and analysis, and graphics for exploration and presentation of messages in data.

Although the collection of forms of census data goes back into antiquity, rulers "were interested in keeping track of their people, money and key events (such as wars and the flooding of the Nile) but little else in the way of quantitative assessment of the world at large" (Scheaffer, 2001, para. 3). The statistical analysis of data is usually traced back to the work of John Graunt (e.g., his 1662 book *Natural and Political Observations*). For example, Graunt concluded that the plague was caused by person-to-person infection rather than the competing theory of "infectious air" based on the pattern of infections through time. Graunt and other "political arithmeticians" from across Western Europe were influenced during the Renaissance by the rise of science based on observation of the natural world. And they "thought as we think today … they reasoned about their data" (Kendall, 1960, p. 448). They estimated, predicted, and learned from the data—they did not just describe or collect facts—and they promoted the notion that state policy should be informed by the use of data rather than by the authority of church and nobility (Porter, 1986). But the political arithmetician's uses of statistics lacked formal methodological techniques for gathering and analyzing data. Methods for sample surveys and census taking were in their infancy well into the nineteenth century (Fienberg, 2014).

Another fundamental thread involved in building modern statistics was the foundation of theories of probability, as laid down by Pascal (1623–1662) and later Bernoulli (1654–1705), which were developed to understand games of chance. The big conceptual steps from that toward the application of probability to inferences from data were taken by Bayes in 1764 and Laplace (1749–1827) by inverting probability analyses, i.e., using knowledge about probabilities of events or data given parameters to make inferences about parameters given data:

> The science that held sway above all others around 1800 was astronomy, and the great mathematicians of the day made their scientific contributions in that area. Legendre (least squares), Gauss (normal theory of errors), and Laplace (least squares and the central limit theorem) all were motivated by problems in astronomy. (Scheaffer, 2001, para. 6)

These ideas were later applied to social data by Quetelet (1796–1874), who with notions such as the "average man" was trying to arrive at general laws governing human action, analogous to the laws of physics. This was after the French Revolution when there was a subtle shift in thinking of statistics as a science of the state with the statists, as they were known, conducting surveys of trade, industrial progress, labor, poverty, education, sanitation, and crime (Porter, 1986).

Another thread in the development of statistics involves statistical graphics (see Friendly, 2008). The first major figure is William Playfair (1759−1823), credited with inventing line charts, bar charts, and the pie chart. Friendly (2008) characterizes the period from 1850 to 1900 as the "golden age of statistical graphics" (p. 2). This is the era of John Snow's famous "dot map" in which he plotted the locations of cholera deaths as dots on a map which then implicated water from the Broad Street pump as a likely cause; of Minard's famous graph showing losses of soldiers in Napoleon's march on Moscow and subsequent retreat; of Florence Nightingale's coxcomb plot used to persuade of the need for better military field hospitals; and of the advent of most of the graphic forms we still use for conveying geographically linked information on maps, including such things as flow diagrams of traffic patterns, of grids of related graphs, of contour plots of three-dimensional tables, population pyramids, scatterplots, and many more.

The Royal Statistical Society began in 1834 as the London Statistical Society (LSS), and the American Statistical Association was formed in 1839 by five men interested in improving the US census (Horton, 2015; Utts, 2015b). Influential founders of the LSS (Pullinger, 2014, pp. 825–827) included Adolphe Quetelet, Charles Babbage (inventor of the computer), and Thomas Malthus (famous for his theories about population growth). The first female LSS member was Florence Nightingale, who joined in 1858 (she also became a member of the ASA, as did Alexander Graham Bell, Herman Hollerith, Andrew Carnegie, and Martin Van Buren). These early members of LSS and ASA were remarkable for representing such a very wide variety of real-world areas of activity (scientific, economic, political, and social) and their influence in society:

> Near the end of the nineteenth century, the roots of a theory of statistics emerge from the work of Francis Galton and Francis Ysidro Edgeworth and from that of Karl Pearson and George Udny Yule somewhat later. These scientists came to statistics from biology, economics, and social science more broadly, and they developed more formal statistical methods that could be used not just within their fields of interest but across the spectrum of the sciences. (Fienberg, 2014, p. 4)

Another wave of activity into the 1920s was initiated by the concerns of William Gosset, reaching its culmination in the insights of Ronald Fisher with the development of experimental design, analysis of variance, maximum likelihood estimation, and refinement of significance testing. This was followed by the collaboration of Egon Pearson and Jerzy Neyman in the 1930s, giving rise to hypothesis testing and confidence intervals. At about the same time came Bruno de Finetti's seminal work on subjective Bayesian inference and Harold Jeffreys's work on "objective" Bayesian inference so that by 1940 we had most of the basics of the theories of the "modern statistics" of the twentieth century. World War II was also a time of great progress as a result of drafting many young, mathematically gifted people into positions where they had to find timely answers to problems related to the war effort. Many of them stayed in the field of statistics swelling the profession. We also draw particular attention to John Tukey's introduction of "exploratory data analysis" in the 1970s; this is an approach to data analysis that involves applying a variety of exploratory techniques, many of them visual, to gain insight into a dataset and uncover underlying structure and exceptions.

Short histories of statistics include Fienberg (2014, Section 3); Scheaffer (2001), who emphasized how mathematicians were funded or employed and the influence this had on what they thought about and developed; and Pfannkuch and Wild (2004) who described the development of statistical thinking. Lengthier accounts are given by Fienberg (1992) and the books by Porter (1986), Stigler (1986, 2016), and Hacking (1990). Key references about the history of statistics education include Vere-Jones (1995), Scheaffer (2001), Holmes (2003), and Forbes (2014); see also Chap. 2.

The current scope and intellectual content of statistics is the result of evolutionary processes involving both slow progress and leaps forward due to the insights of intellectual giants and visionaries, all affected by the intellectual climate of their day and the recognized challenges in the world in which they lived. But it has not reached some fixed and final state. It continues to evolve and grow in response to new challenges and opportunities in the changing environment in which we now live.

1.2.2 Statistical Thinking

Statisticians need to be able to think in several ways: statistically, mathematically, and computationally. The thinking modes used in data analysis differ from those used in working with mathematical derivations, which in turn differ from those used for writing computational code. Although there are very strong internal connections within each of these thinking modes, there are relatively weak connections among them. Here we will concentrate on "statistical thinking" in the sense of the most distinctively statistical parts of the thinking that goes on in solving real-world problems using data.

In statistics, however, we sometimes talk about "solving real-world (or practical) problems" far too loosely. For the general public, "solving a real-world problem" involves taking action so that the problem either goes away or is at least reduced (e.g., unemployment levels are reduced). We need to better distinguish between satisfying "a need to act" and "a need to know." Figuring out how to act to solve a problem will typically require acquiring more knowledge. This is where statistical inquiry can be useful. It addresses "a need to know." So when statisticians talk about solving a real-world problem, we are generally talking about solving a (real-world) *knowledge-deficit* or *understanding-deficit* problem.

1.2.2.1 Statistical Thinking in Statistical Inquiry

Wild and Pfannkuch (1999) investigated the nature of statistical thinking in this sense using available literature, interviews with practicing statisticians, and interviews with students performing statistical-inquiry activities and presented models for different "dimensions" of statistical thinking.

Dimension 1 in Wild and Pfannkuch's description of statistical thinking is the PPDAC model (Fig. 1.1) of the inquiry cycle. The basic PPDAC model was due to and later published by MacKay and Oldford (2000). There are also other essentially equivalent descriptions of the statistical-inquiry cycle. The inquiry cycle has connections with standard descriptions of the scientific method but is more flexible, omitting the latter's strong emphasis on being hypothesis driven and having (scientific) theory formulation as its ultimate objective.

The PPDAC inquiry cycle reminds us of the major steps involved in carrying out a statistical inquiry. It is the setting in which statistical thinking takes place. The initial "P" in PPDAC spotlights the *problem* (or question) crystallization phase. In the early stages, the problem is typically poorly defined. People start with very vague ideas about what their problems are, what they need to understand, and why. The *problem* step is about trying to turn these vague feelings into much more precise informational goals, some very specific questions that should be able to be answered using data. Arriving at useful questions that can realistically be answered using statistical data always involves a lot of hard thinking and often a lot of hard preparatory work. Statistics education research says little about this, but the PhD thesis of Arnold (2013) makes a very good start.

The *plan* step is then about deciding what people/objects/entities to obtain data on, what things we should "measure," and how we are going to do all of this. The *data* step is about data acquisition, storage, and wrangling (reorganizing the data using various transformations, merging data from different sources, and cleansing the data in preparation for analysis). The *analysis* step which follows and the *conclusions* step are about making sense of it all and then abstracting and communicating what has been learned. There is always a back-and-forth involving doing analysis, tentatively forming conclusions, and doing more analysis. In fact there is back-and-forth between the major steps whenever something new gets learned in a subsequent step that leads to modifying an earlier decision (Konold & Pollatsek, 2002).

Any substantial learning from data involves extrapolating from what you can see in the data you have to how it might relate to some wider universe. PPDAC focuses on data gathered for a purpose using planned processes, processes that are chosen on statistical grounds to justify certain types of extrapolation. Much of the current buzz about the widespread availability and potential of data (including "big data") relates to exploiting opportunistic (happenstance or "found") data—data that just happen to be available in electronic form because they have accumulated for other reasons, such as the result of the administrative processes of business or government, audit trails of internet activity, or billing data from medical procedures:

> In a very real sense, we have walked into the theatre half way through the movie and have then to pick up the story. … For opportunistic data there is no extrapolation that is justified by a data-collection process specifically designed to facilitate that extrapolation. The best we can do is to try to forensically reconstruct what this data is and how it came to be (its 'provenance'). What entities were 'measures' taken on? What measures have been employed and how? By what processes did some things get to be recorded and others not? What distortions might this cause? It is all about trying to gauge the extent to which we can

> generalize from patterns in the data to the way we think it will be in populations or processes that we care about. (Wild, 2017, p. 34)

In particular, we are on the lookout for biases that could lead us to false conclusions.

Dimension 2 of Wild and Pfannkuch's model lists types of thinking, broken down into general types and types fundamental to statistics. The general types are *strategic*, *seeking explanations*, constructing and *using models*, and *applying techniques* (solving problems by mapping them on to problem archetypes). The types fundamental to statistics listed are *recognition of the need for data*, *transnumeration* (changing data representations in search of those that trigger understanding), *consideration of variation* and its sources, *reasoning using statistical models*, and *integrating the statistical and the contextual* (information, knowledge, conceptions). Something that is not highlighted here is the inductive nature of statistical inference—extrapolation from data on a part to reach conclusions about a whole (wider reality).

Dimension 3 is the *interrogative cycle*, a continually operating high-frequency cycle of *generating* (possible informational requirements, explanations, or plans of attack), *seeking* (information and ideas), *interpreting* these, *criticizing* them against reference points, and *judging* whether to accept, reject, or tentatively entertain them. Grolemund and Wickham (2014) dig much deeper into this dimension bringing in important ideas from the cognitive literature.

Dimension 4 consists of a list of personal qualities, or *dispositions*, successful practitioners bring to their problem solving: *skepticism*, *imagination*, *curiosity and awareness*, a *propensity to seek deeper meaning*, *being logical*, *engagement*, and *perseverance*. This is amplified in Hahn and Doganaksoy's chapter "Characteristics of Successful Statisticians" (2012, Chapter 6).

1.2.2.2 Statistical Thinking for Beginners

Although it only scratches the surface, the above still underscores the richness and complexity of thinking involved in real-world statistical problem solving and provides a useful set of reference points against which researchers and teachers can triangulate educational experiences ("Where is … being addressed?"). It is, however, far too complex for most students, particularly beginners. In discussing Wild and Pfannkuch, Moore (1999) asked, "What shall we teach beginners?" He suggested:

> … we can start by mapping more detailed structures for the 'Data, Analysis, Conclusions' portion of the investigative cycle, that is, for conceptual content currently central to elementary instruction. Here is an example of such a structure:
>
> When you first examine a set of data, (1) begin by graphing the data and interpreting what you see; (2) look for overall patterns and for striking deviations from those patterns, and seek explanations in the problem context; (3) based on examination of the data, choose

> appropriate numerical descriptions of specific aspects; (4) if the overall pattern is sufficiently regular, seek a compact mathematical model for that pattern. (Moore, 1999, p. 251)

Moore (1998) offered the following for basic critique, which complements his 1999 list of strategies with "Data beat anecdotes" and the largely metacognitive questions, "Is this the right question? Does the answer make sense? Can you read a graph? Do you have filters for quantitative nonsense?" (p. 1258).

There are great advantages in short, snappy lists as starting points. Chance's (2002) seven habits (p. 4) bring in much of Moore's lists, and the section headings are even "snappier": "Start from the beginning. Understand the statistical process as a whole. Always be skeptical. Think about the variables involved. Always relate the data to the context. Understand (and believe) the relevance of statistics. Think beyond the textbook." Grolemund and Wickham (2014, Section 5) give similar lists for more advanced students. Brown and Kass (2009) state, "when faced with a problem statement and a set of data, naïve students immediately tried to find a suitable statistical technique (e.g., chi-squared test, *t*-test), whereas the experts began by identifying the scientific question" (p. 123). They highlighted three "principles of statistical thinking":

1. Statistical models of regularity and variability in data may be used to express knowledge and uncertainty about a signal in the presence of noise, via inductive reasoning. (p. 109)
2. Statistical methods may be analyzed to determine how well they are likely to perform. (p. 109)
3. Computational considerations help determine the way statistical problems are formalized. (p. 122)

We conclude with the very specialized definition of Snee (1990), which is widely used in quality improvement for business and organizations:

> I define statistical thinking as thought processes, which recognize that variation is all around us and present in everything we do, all work is a series of interconnected processes, and identifying, characterizing, quantifying, controlling, and reducing variation provide opportunities for improvement. (p. 118)

1.2.3 Relationship with Mathematics

Although definitions that characterize statistics as a branch of mathematics still linger in some dictionaries, the separate and distinct nature of statistics as a discipline is now established. "Statistical thinking," as Moore (1998) said, "is a general, fundamental, and independent mode of reasoning about data, variation, and chance" (p. 1257). "Statistics at its best provides methodology for dealing empirically with complicated and uncertain information, in a way that is both useful and scientifically valid" (Chambers, 1993, p. 184).

"Statistics is a methodological discipline. It exists not for itself but rather to offer to other fields of study a coherent set of ideas and tools for dealing with data" (Cobb & Moore, 1997, p. 801). To accomplish those ends it presses into service any tools that are of help. Mathematics contains many very useful tools (as does computing). Just as physics attempts to understand the physical universe and presses mathematics into service wherever it can help, so too statistics attempts to turn data into real-world insights and presses mathematics into service wherever it can help. And whereas in mathematics, mathematical structures can exist and be of enormous interest for their own sake, in statistics, mathematical structures are merely a means to an end (see also Box, 1990, paragraph 2; De Veaux & Velleman, 2008). A consequence is, adapting a famous quotation from John Tukey, whereas a mathematician prefers an exact answer to an approximate question, an applied statistician prefers an approximate answer to an exact question.

The focus of the discipline of statistics, and in particular the role of *context*, is also distinct. "Statistics is not just about the methodology in a particular application domain; it also focuses on how to go from the particular to the general and back to the particular again" (Fienberg, 2014, p. 6):

> Although mathematicians often rely on applied context both for motivation and as a source of problems for research, the ultimate focus in mathematical thinking is on abstract patterns: the context is part of the irrelevant detail that must be boiled off over the flame of abstraction in order to reveal the previously hidden crystal of pure structure. *In mathematics, context obscures structure*. Like mathematicians, data analysts also look for patterns, but ultimately, in data analysis, whether the patterns have meaning, and whether they have any value, depends on how the threads of those patterns interweave with the complementary threads of the story line. *In data analysis, context provides meaning*. (Cobb & Moore, 1997, p. 803; our emphasis)

There is a constant "interplay between pattern and context" (Cobb & Moore, 1997). As for statistical investigations for real-world problems, the ultimate learning is new knowledge about the context domain—we have gone "from the particular to the general" (to enable us to use methods stored in our statistical repository) "and back to the particular again" (to extract the real-world learnings).

We will now turn our attention to the role of theory in statistics. When most statisticians speak of *statistical theory*, they are thinking of mathematical theories comprising "statistical models" and principled ways of reasoning with and drawing conclusions using such models. *Statistical models*, which play a core role in most analyses, are mathematical models that include chance or random elements incorporated in probability theory terms. Perhaps the simplest example of such a model is $y = \mu + \varepsilon$ where we think in terms of y being an attempt to measure a quantity of interest μ in the process of which we incur a random error ε (which might be modeled as having a normal distribution, say). In the simple linear model, $y = \beta_0 + \beta_1 x + \varepsilon$, the mean value of y depends linearly on the value of an explanatory variable x rather than being a constant. Random terms (cf. ε) model the unpredictable part of a process and give us ways of incorporating and working with uncertainties. "Statistical theory" in this sense is largely synonymous with "mathematical statistics."

Probability theory is a body of mathematical theory which was originally motivated by games of chance. More recently it has been motivated much more by the needs of statistical modeling, which takes abstracted ideas about randomness, forms mathematical structures that encode these ideas and makes deductions about the behavior of these structures. Statistical modelers use these structures as some of the building blocks that they can use in constructing their models, as with the random error term in the very simple model above. Recent work by Pfannkuch et al. (2016) draws on interviews with stochastic modeling practitioners to explore probability modeling from a statistical education perspective. The paper offers a new set of (conceptual) models of this activity. Their SWAMTU model is basically a cycle (with some feedback). It has nodes *problem Situation* → *Want (to know)* → *Assumptions* → *Model* → *Test* → *Use*. Models are always derived from a set of mathematical assumptions so that assumption checking against data is, or should be, a core part of their construction and use. As well as being used in statistical analysis, they are commonly used to try to answer "what if" questions (e.g., "What would happen if the supermarket added another checkout operator?"). Although there is much that is distinct about statistical problem solving, there is also much that is in common with mathematical problem solving so that statistics education researchers can learn a lot from work in mathematics education research and classic works such as Schoenfeld (1985).

When most statisticians hear "theory," they think "statistical theory" as described above, mathematical theories that underpin many important practices in the analysis and plan stages of PPDAC. But "theory" is also applicable *whenever* we form abstracted or generalized explanations of how things work. Consequently, there is also theory about other elements of PPDAC, often described using tools like flow charts and concept maps (cf. Fig. 1.1). For example, Grolemund and Wickham (2014) propose a theoretical model for the data analysis process by comparing it to the cognitive process of the human mind called "sensemaking" involving updating schemas (mental models of the world) in light of new information.

In recent years there has also been a shift in the "balance of power" from overtly mathematical approaches to data analysis toward computationally intensive approaches (e.g., using computer simulation-based approaches including bootstrapping and randomization tests, flexible trend smoothers, and classification algorithms). Here the underlying models make much weaker assumptions and cannot be described in terms of simple equations. So, although "the practice of statistics requires mathematics for the development of its underlying theory, statistics is distinct from mathematics and requires many nonmathematical skills" (American Statistical Association Undergraduate Guidelines Workgroup, 2014, p. 8). These skills (required also by many other disciplines) include basic scientific thinking, computational/algorithmic thinking, graphical/visualization thinking (Nolan & Perrett, 2016), and communication skills.

So, "… how is it then that statistics came to be seen as a branch of mathematics? It makes no more sense to us than considering chemical engineering as a branch of mathematics" (Madigan & Gelman, 2009, p. 114). The large majority of senior statisticians of the last half century began their academic careers as mathematics majors. Originally, computational capabilities were extremely limited, and mathematical solutions to simplified problems and mathematical approximations were hugely important (Cobb, 2015). The academic statisticians also worked in environments where the reward systems overwhelmingly favored mathematical developments. The wake-up call from "big data" and "data science" is helping nudge statistics back toward its earlier, and much more holistic, roots in broad scientific inference (Breiman, 2001).

1.3 Why Learning Statistics Is More Important Than Ever

In today's data-rich world, all educated people need to understand statistical ideas and conclusions, to enrich both their professional and personal lives. The widespread availability of interesting and complex data sets and increasingly easy access to user-friendly visualization and analysis software mean that anyone can play with data to ask and answer interesting questions. For example, Wild's Visual Inference Tools (https://www.stat.auckland.ac.nz/~wild/VIT/) and iNZight software (https://www.stat.auckland.ac.nz/~wild/iNZight) allow anyone to explore data sets of their own choosing. The CODAP (Common Online Data Analysis Platform, https://concord.org/projects/codap) provides a straightforward platform for web-based data analysis, as does iNZight Lite (http://lite.docker.stat.auckland.ac.nz/), and commercial solutions created by TuvaLabs (https://tuvalabs.com), and Tableau (https://www.tableau.com).

Statistical methods are used in almost all knowledge areas and increasingly are used by businesses, governments, health practitioners, other professionals, and individuals to make better decisions. Conclusions and advice based on statistical methods abound in the media. Some of the thinking used for decision-making based on quantitative data carries over into decision-making involving uncertainty in daily life even when quantitative data are not available. For these reasons, probably no academic subject is more useful to both working professionals and informed citizens on a daily basis than statistics.

The rapid development of data science and expansion of choices for what to teach in statistics courses provides challenges for statistics educators in determining learning goals, and opportunities for statistics education researchers to explore what instructional methods can best achieve those goals. For example, in articles that appeared almost simultaneously, both Cobb (2015) and Ridgway (2015) argued that we need a major overhaul of the university statistics curriculum, particularly the introductory course and the undergraduate curriculum. Similar arguments can be made for greater inclusion of data skills at the primary and secondary school levels.

To ignore the impact of the widespread availability of data and user-friendly software to play with data would lead to marginalization of statistics within the expanding world of data science.

An important and engaging component of the current data revolution is the existence of "data traces" of people's day to day activities, captured in social networks, personal logging devices, and environmental sensors. Teaching students how to examine their own personal data in constructive ways can enhance the attractiveness of learning data skills. These advances offer people new ways to become agents and advocates empowered to use data to improve the world around them in relation to situations of special relevance to their own lives (Wilkerson, 2017, personal communication).

In this section we provide some additional motivation for why everyone would benefit from studying statistics and some examples of what could be useful for various constituencies. Statistics educators could then make sure to emphasize the usefulness of statistics when they teach, targeted to their respective audiences. Statistics education researchers study how we can teach students to use statistical reasoning throughout their lives to ask and answer questions relevant to them.

The ideal type and amount of statistical knowledge and competencies needed by an individual depends on whether the person will eventually be working with data as a professional researcher (a *producer* of statistical studies), interpreting statistical results for others (a *professional user* of statistics), or simply needing to understand how to use data and interpret statistical information in life (an *educated consumer* of data and statistics). Professional users include health workers (who need to understand results of medical studies and translate them into information for patients), financial advisors (who need to understand trends and variability in economic data), and politicians (who need to understand scientific data as it relates to public policy, as well as how to conduct and understand surveys and polls). Producers of statistical studies are likely to take several courses in statistical methods, and will not be the focus of this chapter (see Sect. 1.4 for more discussion). Educated consumers include pretty much everyone else in a modern society. They need to understand how and what valid conclusions can be made from statistical studies and how statistical thinking can be used as a tool for answering questions and making decisions, with or without quantitative data.

1.3.1 What Professional Users of Statistics Need to Know

Many professionals do not need to know how to carry out their own research studies, but they do need to know how to interpret and question results of statistical studies and explain them to patients and customers. In business applications, professionals such as marketing managers may need to understand the results generated by statisticians within their own companies. In this section we provide a few examples of why professional users of statistics need to understand basic statistical ideas beyond what is needed for the general consumer. Commonly used statistical methods differ somewhat across disciplines, but there are some basic ideas that apply to almost all of them. One of the most fundamental concepts is the importance of variability and a distribution of values. The first example illustrates how that concept is important for financial advisors and their clients.

1.3.2 Example 1: How Much Should You Save? Income and Expense Volatility

In 2015 the financial giant JP Morgan Chase announced the establishment of a research institute to utilize the massive and proprietary set of financial data it owns to answer questions about consumer finances. The inaugural report, published in May 2015 (Farrell & Greig, 2015), examined financial data for 100,000 individuals randomly selected from a 2.5 million person subset of Chase customers who met specific criteria for bank and credit card use. One of the most important and publicized findings (e.g., Applebaum, 2015) was that household income and expenditures both vary widely from month to month and not necessarily in the same direction. For instance, the report stated that "41% of individuals experienced fluctuations in income of more than 30% on a month-to-month basis" (p. 8) and "a full 60% of people experienced average monthly changes in consumption of greater than 30%" (p. 9). Additionally, the report found that changes in income and consumption don't occur in tandem, so it isn't that consumers are spending more in months when they earn more. Why is this important information? Financial planners routinely advise clients to have accessible savings equivalent to 3–6 months of income. But in some cases, that may not be enough because of the volatility in both income and expenditures and the possibility that they can occur in opposite directions. One of three main findings of the report was "the typical individual did not have a sufficient financial buffer to weather the degree of income and consumption volatility that we observed in our data." (p. 15).

A concept related to variability is that few individuals are "typical." In the previous example, individual consumers should know whether the warnings in the report are likely to apply to them based on how secure their income is, what major expenditures they are likely to encounter, and what resources they have to weather financial storms. Knowing that the "typical" or "average" consumer needs to stash away 3–6 months of income as savings could lead to questions about how each individual's circumstances might lead to recommendations that differ from that advice.

Baldi and Utts (2015) discussed topics and examples that are important for future health practitioners to learn in their introductory, and possibly only, statistics course. One central concept is that of natural variability and the role it plays in defining disease and "abnormal" health measurements. The next example, adapted from Baldi and Utts, illustrates how knowledge about variability and a distribution of

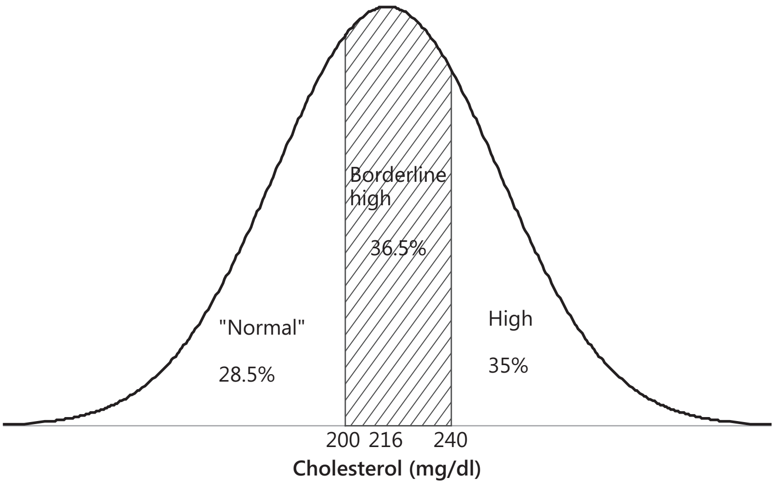


Fig. 1.2 Cholesterol values for women aged 45–59; mean ≈ 216 mg/dL (standard deviation ≈ 42 mg/dL)

values can help physicians and their patients put medical test results in perspective.

1.3.3 Example 2: High Cholesterol for (Almost) All

Medical guidelines routinely change based on research results, and statistical studies often lead pharmaceutical companies and regulatory agencies to change their advice to medical practitioners about what constitutes pathology. According to the United States National Institutes of Health, high cholesterol is defined as having total blood cholesterol of 240 mg/dL or above, and elevated or borderline high cholesterol is defined as between 200 and 240 mg/dL (http://www.nhlbi.nih.gov/health/health-topics/topics/hbc). Suppose you are diagnosed with high or borderline high cholesterol and your doctor recommends that you take statin drugs to lower it. You might be interested in knowing what percentage of the population is in the same situation. Using data from the World Health Organization (Lawes, Vander Hoorn, Law, & Rodgers, 2004), we can model the total cholesterol levels of women aged 45–59 years old in the United States using a normal distribution with mean of about 216 mg/dL and standard deviation of about 42 mg/dL. Figure 1.2 illustrates this distribution. As shown in the figure, about 35% of women in this age group have high cholesterol, and an additional 36.5% have borderline high values. That means that only about 28.5% do not have this problem! Should more than 70% of middle-aged women be taking statin drugs? Given that there are risks associated with high cholesterol and side effects associated with taking statin drugs, this is a discussion for individuals to have with their doctors. But it would be helpful for both of them to understand that more than a majority of the population fall into these cholesterol risk categories. Additional statistical reasoning tools (beyond the scope of this example) are needed to help physicians and consumers understand the trade-off in risks associated with taking statin drugs or not. But this example illustrates the importance of understanding the concept of a distribution of values, and how it relates to decisions individuals need to make.

1.3.4 What Educated Consumers of Data and Statistics Need to Know

Most students who take a statistics course will never use formal statistical procedures in their professional lives, but quite often they will encounter situations for which they could utilize data and statistical information in their personal lives to make informed decisions. Teachers of introductory statistics should provide instruction that will help people utilize that information. In the papers by Ridgway (2015), Baldi and Utts (2015), Utts (2003), and Utts (2010), more than a dozen important topics are described, with multiple examples. A few of them were covered in the previous section on what professional users of statistics need to know. Here, we list more of them and explain why they are important for consumers. Examples to illustrate each of these can be found in the references mentioned here, as well as in textbooks with a focus on statistical literacy such as Cohn and Cope (2011), Hand (2014), Moore and Notz (2016), and Utts (2015a). A resource for current examples is the website http://www.stats.org, a joint venture of the American Statistical Association and Sense About Science USA.

1.3.4.1 Unwarranted Conclusions Based on Observational Studies

The media have improved in the interpretation of observational studies in recent years, but reports implying causation based on observational studies are still quite common. Citizens should learn to recognize observational studies and know that cause and effect conclusions cannot be made based on them. Here are some recent examples of misleading headlines based on observational studies or meta-analyses of them:

- “6 cups a day? Coffee lovers less likely to die, study finds” (NBC News, 2012)
- “Citrus fruits lower women’s stroke risk” (Live Science, 2012)
- “Walk faster and you just might live longer” (NBC News, 2011)

In many cases those conducting the research cautioned against making a causal conclusion, but the headlines are what people read and remember. Students and citizens should always question whether a causal conclusion is warranted and should know how to answer that question.

1.3.4.2 Statistical Significance Versus Practical Importance

Ideally, university students would learn the correct interpretation of *p*-values in an introductory statistics course. (See Nuzzo, 2014, for a nontechnical explanation of *p*-values.) But the most common misinterpretation, that the *p*-value measures the probability that chance alone can explain observed results, is difficult to overcome. So students at least should learn the importance of distinguishing between statistical significance and practical importance.

For example, in the previously referenced article, Nuzzo (2014) discusses a study claiming that those who meet online have happier marriages ($p < 0.001$) and lower divorce rates ($p < 0.002$) than those who meet offline. The *p*-values of less than 0.001 and 0.002 look impressive, but on a 7-point scale, the average "happiness" ratings for the two groups were 5.48 and 5.64, and the divorce rates were 7.67% and 5.96%. These differences are of little practical importance. The small *p*-values resulted from the large sample size of over 19,000 people.

The easiest way to illustrate the difference between statistical significance and practical importance is to look at a variety of studies that have small *p*-values, such as the one just described, and then look at a confidence interval for the population parameter in each case. Although *p*-values are often used in situations that don't involve an easily interpretable parameter (for which a confidence interval could be computed), showing examples for which there is an interpretable parameter will make the point about statistical versus practical significance and especially about the importance of sample size.

In 2016 the American Statistical Association took the unusual step of publishing a statement on the use and misuse of *p*-values (Wasserstein & Lazar, 2016) along with commentaries by numerous statisticians. Within a year of publication, the statement was viewed nearly 200,000 times, indicating widespread interest in learning more about how to use and interpret *p*-values.

Aspects such as the size of the study and how that impacts the *p*-value are important if students are to understand the distinction between statistical significance and practical importance. Introducing the concept of "effect size" can help make this point, as illustrated in the next section.

1.3.4.3 The Difference Between No Effect and No Statistically Significant Effect

Studies that find no statistically significant effect may never make it into publication because they are not thought to have found anything of interest and are not newsworthy. But often media reports will mention an "unsuccessful replication" of an earlier study and report it as if it contradicts the earlier result. Consumers should understand that there are multiple reasons for this apparent contradiction, and they do not all imply a real contradiction.

Consider the following scenario, called the "hypothesis testing paradox" (Utts & Heckard, 2015, p. 542). A researcher conducts a *t*-test for a population mean based

Table 1.1 Hypothetical example of the relationships among sample size, test statistic, and p-value

Study	n	Effect size $(\bar{x} - \mu_0)/s$	Test statistic t	p-value
1	100	0.25	2.50	0.014
2	25	0.25	1.25	0.22
Combined	125	0.25	2.80	0.006

The same effect size yields a statistically significant result for a larger study

on a sample of $n = 100$ observations, obtaining a result of $t = 2.50$ and $p = 0.014$, so the null hypothesis is rejected. The experimenter decides to repeat the experiment with $n = 25$ observations to verify the result but finds disappointingly that the result is $t = 1.25$, $p = 0.22$, and so she cannot reject the null hypothesis. The effect seems to have disappeared. To salvage the situation, she decides to combine the data, so now $n = 125$. Based on the combined data, $t = 2.80$, p-value = 0.006! How could a second study that seemed to diminish the statistically significant result of the first study somehow make the result stronger when combined with the first study?

The paradox is that the second study alone did not replicate the finding of statistical significance, but when combined with the first study, the effect seems even stronger than the first study alone, with the p-value going from 0.014 to 0.006. The problem of course is that the test statistic and p-value depend on the sample size. In fact in this example, the effect size (measured as $(\bar{x} - \mu_0)/s$) is the same in both studies. It is the sample size that creates the difference in t and the p-value. See Table 1.1 for a numerical explanation.

1.3.4.4 Sources of Potential Bias in Studies and Surveys and the Population to Which Results Apply

As fewer households maintain landline telephones and caller ID makes it easy to ignore calls from unfamiliar numbers, it is becoming increasingly difficult to get representative samples for surveys and other statistical studies. Consequently, in many cases the results of surveys and other studies may not reflect the population of interest. According to Silver (2014), "Even polls that make every effort to contact a representative sample of voters now get no more than 10 percent to complete their surveys—down from about 35 percent in the 1990s" (para. 1).

Lack of response is just one of many sources of bias that can enter into surveys and other studies. Other sources of bias include poor and/or intentionally biased wording of questions, the order in which questions are asked, who is asking, and whether the topic is one for which people are inclined to lie (for instance, to appear to conform to social norms). When reading results of surveys, it's important to know exactly what was asked; who asked; whether questions were asked in person, by mail, by phone, or online; and whether special interest groups were involved in any way that could affect the results. Cohn and Cope (2011) provide a detailed list of questions journalists should ask when covering these kinds of studies, but the list is relevant to anyone interested in learning more about how to detect bias in survey results. Questions cover things like finding out exactly how the questions were worded, how the respondents were selected, what percent of those contacted responded, whether the survey was funded by a special interest group, and whether the possible response choices were chosen to favor a certain viewpoint.

1.3.4.5 Multiple Testing and Selective Reporting

Almost all studies measure multiple explanatory and/or response variables and therefore conduct multiple statistical tests to find out which variables are related. It is very common for the media to report the findings that happen to be statistically significant without mentioning that they were part of a larger study. If the original research report did not correct the reported results for multiple testing, the statistically significant findings could easily be spurious. Given that 20 independent tests with true null hypotheses are expected to yield one with statistical significance, it is not surprising that false claims make their way into the media. Students should learn to ask how many different tests were conducted in a study and whether the statistical results were adjusted for multiple testing. Ioannidis (2005) illustrates this problem and related issues with many examples in his article "Why most published research findings are false."

1.3.4.6 Interpreting Relative Risk, Absolute Risk, Personal Risk, and Risk Trade-Offs

Knowing how to think about risk can help consumers in many ways. Education research has shown that teaching risk and relative risk in terms of frequencies instead of probabilities will make them easier for most students to understand, and giving baseline risks in addition to relative risks will allow people to make more informed decisions (Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2008), particularly when the base rate is very low. For example, students understand the idea that 3 out of 1000 people may die from a certain treatment more readily than they understand that 0.003 or 0.3% of those treated may die. Saying that a certain behavior doubles your probability (or risk) of cancer from 0.003 to 0.006 is not easy for most people to understand, but saying that the behavior increases the *number* of cases of cancer in people similar to them from 3 in 1000 to 6 in 1000 is much easier to understand. And reporting frequencies instead of proportions makes the role of baseline risk much more clear. Most people can understand that an increase from 3 in 1000 to 6 in 1000 is different than an increase from 30 in 1000 to 60 in 1000 and will immediately recognize what the baseline risk is in each case.

As an example, Utts (Utts, 2015a, p. 258) describes a study showing that the risk of esophageal cancer in men who drink alcohol daily is about three times the risk for men who don't drink (the baseline risk). If you are a male who drinks daily, how much should this concern you? According to statistics from the United States

National Cancer Institute, the annual incidence of esophageal cancer in men is about 7.7 cases per 100,000 men. That statistic includes both drinkers and nondrinkers, so let's guess that the baseline risk, for nondrinkers, is about 5 in 100,000. That would tell us that the risk for daily drinkers is about 15 per 100,000. This is not nearly as worrisome as it would be if the tripled risk meant a change of 5 per hundred to 15 per hundred. Knowing the baseline risk is important in deciding how much of a concern a particular relative risk might be.

Another important feature of risk to explain to students is that changing one behavior to avoid risk may lead to increasing risk of a different outcome. For instance, taking drugs to reduce blood pressure or cholesterol may increase the risk of other medical problems. Having a mammogram to reduce the risk of undetected breast cancer may increase the risk of the effects of radiation, or add a psychological risk of having a false positive result, and the accompanying stress. A striking example by Gigerenzer et al. (2008), described by Utts (2010), illustrated how a media scare associated with birth control pills in the United Kingdom resulted in reduced use of the pills but led to large increases in abortions and teen pregnancies, which had much higher risks than the use of the pills would have had. Educated citizens should understand how to view behavior changes that reduce risk in the broader context of risk trade-offs.

1.3.4.7 Conditional Probability and "Confusion of the Inverse"

Psychologists know that people have very poor intuition about probability. One example is called "confusion of the inverse," in which people confuse conditional probability in one direction with conditional probability in the other direction. A classic example is confusing the probability of a positive test result given that you have a disease with the probability of having the disease, given a positive test result. The two probabilities can of course be vastly different, and this confusion has undoubtedly led to much unnecessary angst in people who receive a false positive medical test result. In a classic study that helped psychologists understand this phenomenon, Eddy (1982) showed that physicians gave estimates of the probability of having breast cancer given a positive mammogram that were ten times too high, close to 0.75 when the actual probability was 0.075.

Another example of this confusion is in the courtroom, where the probability of guilt given particular evidence is not the same as the probability of that evidence, given that the person is guilty. Again these can be vastly different, and juries need to recognize the difference. For example, suppose size 13 shoe prints are found at the scene of a crime. Then it is quite likely that, given a person is guilty of that crime, the person wears size 13 shoes. But it is not at all likely that (without other evidence) a person is guilty of the crime, given that he wears size 13 shoes.

Similar to explanations of relative risk (as in the previous example), conditional probabilities are easier to understand using frequencies instead of proportions or probabilities. One of the easiest ways to illustrate conditional probabilities in both directions, and how they differ, is through the use of a "hypothetical hundred

Table 1.2 The probability of having the disease given a positive test

	Test positive	Test negative	Total
Have disease	297	3	300
Do not have disease	997	98,703	99,700
Total	1294	98,706	100,000

thousand" table. Suppose a disease occurs on average in three out of 1000 people in the population and that a test for the disease is 99% accurate when someone has the disease (the test says they do) and when they don't have the disease (the test says they do not). What is the probability that a person has the disease, given that a test is positive? A table of a hypothetical 100,000 people can be constructed as follows (Table 1.2). First, fill in the row totals. If three out of 1000 have the disease, then 300 out of 100,000 have it, and 99,700 do not. Next fill in the cells in each row for positive and negative tests. 99% of the 300 with the disease is 297 people, which shows that 297 test positive and three test negative. Similarly, of the 99,700 without the disease, 1%, or 997 test positive, and the remaining 98,703 would test negative. So the hypothetical 100,000 people would distribute as shown in the table, and the probability of having the disease given a positive test is easily seen to be 297/1294 = 0.23.

See Utts (2010) for another example, as well as for other ways that psychological influences can affect probability estimates and interpretation.

1.3.5 What Decision Makers Need to Know

Statistical ideas and methods provide many tools for making decisions in life, especially decisions that involve trade-offs. Previously we discussed how competing risks often need to be taken into account when making decisions. We now discuss some other ways in which statistical ideas can help with making decisions when trade-offs are involved.

1.3.5.1 The Importance of Expected Values to Make Better Decisions

Insurance companies, casinos, lottery agencies, and sellers of extended warranties all rely on expected values and can exploit consumers who do not understand them. In all of those cases, consumers overall are the losers, but sometimes the protection (as with insurance and extended warranties) is worth the loss. The important point for consumers is to understand how to figure out when that is true. As an example, if you buy an appliance, should you buy the extended warranty? If you have sufficient income or financial reserves to have it fixed or replaced, the answer is probably not. In the long run, you will lose money if you often buy insurance and extended warranties. But an individual consumer does not get the benefit of "the long run" in something like the purchase of a house or car, so in those cases it may be worth having the insurance in the (small probability) event of a disaster. Also, if you are a real klutz and tend to break things, you might actually come out ahead with certain extended warranties. Students should understand these issues so they can make informed choices.

1.3.5.2 Example: Should You Pay in Advance?

Here is a simple example of where knowledge of expected value could be useful in making a decision (Utts, 2015a, Exercise 17.29). Suppose you are planning to stay at a hotel a month from now but are not 100% sure you will take the trip. The hotel offers two payment choices. You can pay an "advance purchase" price of $85 now, nonrefundable. Or, you can reserve a room and pay $100 when you go, but not pay anything if you decide not to go. Which choice should you make? With the advance purchase, the "expected value" of what you will pay is $85, because the probability is 1.0 that you will pay that amount. Define p to be the probability that you will take the trip. If you don't use the advance purchase option, the expected value for what you will pay is $(\$100)(p) + (\$0)(1 - p) = \$100p$. Note that $\$100p$ is less than $85 if p is less than 0.85. So if you think the probability of taking the trip is less than 0.85, the advance purchase is not a good idea, but if you think the probability is higher than 0.85, the expected value is lower by taking advantage of the advance purchase.

1.3.6 Using the Hypothesis Testing Framework to Make Decisions

In addition to the technical aspects of statistical studies, the reasoning used for hypothesis testing can be useful in making decisions even without quantitative data. Consider the following example from the life of one of the authors.

Table 1.3 Did the dog eat the chocolate?

	Decision	
Hypothesis	*Dog did not eat the chocolate*	*Dog did eat the chocolate*
Null: Dog did not eat the chocolate	No trip to veterinarian; OK	Type 1 error: Go to vet; dog has stomach pumped needlessly
Alternative: Dog did eat the chocolate	Type 2 error: Do not go to vet. Dog could die	Go to vet; thank goodness you had stomach pumped

The empty wrapper for a chocolate bar is sitting on a table, carelessly left where your dog could find it. The person involved cannot remember whether they ate the whole chocolate bar or left half of it exposed on the table. You fear that the dog consumed the remainder of the chocolate, an indulgence that could be fatal to the dog. Should you rush to the veterinarian to have your dog's stomach pumped?

We can think about this decision in the framework of hypothesis testing and look at the equivalent of type 1 and type 2 errors when considering the decision, as shown in Table 1.3. The actual decision will depend on how likely you think the two hypotheses are, but illustrating the possible choices and their consequences can be informative and helpful in making a decision.

In this example, if there was even a relatively small chance that the dog ate the chocolate, most dog owners would be likely to take the dog to the vet for an evaluation. In general, the decision would be based on the seriousness of the consequences of the two types of errors. Laying the choices out in this kind of table makes those consequences clearer.

1.3.7 Final Remarks About the Importance of Statistics

It is hard to predict the future of data science and statistics, as resources become available that allow easy access to data and methods for visualizing and analyzing them. As eminent statistician Brad Efron noted, "Those who ignore statistics are condemned to reinvent it" (attributed to Efron by Friedman, 2001, p. 6), and as Wild (2015) notes, "their ignorance can do real damage in the meantime" (p. 1). Statistics educators have a grave responsibility and an exciting opportunity to make sure that everyone learns how useful statistics can be. Statistics education researchers have their work cut out for them in figuring out how best to convey these ideas in ways that are useful and that will allow students to make better decisions using data.

1.4 Where Statistics Is Heading

The previous sections have described the nature of statistics and the importance of statistical education. In this section we discuss current developments and make suggestions about where the field is heading.

1.4.1 An Exciting Time to be a Statistician

This is an exciting time to be a statistician. Interest in the discipline of statistics and the analysis of data is booming. The amount of information collected in our increasingly data-centered society is staggering. Statistical expertise is more valuable than ever, with society and employers clamoring for graduates who can blend knowledge of statistics, data management, computing, and visualization to help make better decisions. But with the opportunities afforded by this rich information come threats for statistics as a discipline. What does the future hold? What do we need to be addressing to ensure that students are developing the statistical skills, knowledge, and competencies they need for their lives and careers?

Speaking of the recent availability of a vast flood of data is not hyperbole. George Lee of Goldman Sachs estimates that 90% of the world's data have been created in the last 2 years (http://www.goldmansachs.com/our-thinking/pages/big-data.html). The 2013 Future of Statistics (2013) report enumerates examples such as astronomy, where new telescopes will generate a petabyte of data each day and commercial databases at social media companies such as Facebook, which generate more than 500 terabytes per day. United States President Barack Obama signed an Open Data Executive Order in 2013 (https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-) that called for data on health, energy, education, safety, finance, and global development to be made machine accessible to "generate new products and services, build businesses, and create jobs," and this has led to increased access to sophisticated and detailed information.

These increasingly diverse data are being used to make decisions in all realms of society. Consider the theme for the American Association for the Advancement of Science (AAAS) annual meeting in 2015 (Innovations, Information, and Imaging) that focused on the ways that science and technology are being transformed by new ways to collect and use information, with progress increasingly driven by the ability to organize, visualize, and analyze data (AAAS, 2015).

Planning a trip to New York City (NYC)? It's straightforward to download and analyze data on all commercial flights in the United States since 1987 (180 million records, http://www.amherst.edu/~nhorton/precursors), 14 million taxi rides in 2013 (http://www.andresmh.com/nyctaxitrips/) and over a billion records for 2009–2015 (http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml), millions of Citi Bike rentals (http://www.citibikenyc.com/system-data), and restaurant violations (http://www.nyc.gov/html/doh/html/services/restaurant-inspection.shtml). Useful information of this type is widely available for many other large cities, governments, and domains.

It's worth noting that although much is said about "big data," none of these NYC examples qualify. A typical definition of "big data" requires datasets with sizes that are difficult to process in a timely fashion using a typical workflow. This definition references the 3Vs model (http://www.gartner.com/newsroom/id/1731916) of volume, velocity, and variety. The NYC examples have a modest *volume* but low *velocity* and modest *variety* (forms of data not easily stored in a rectangular array). Though issues of "big data" are important, many more challenges and opportunities are available for smaller scale information.

The development of new and easier to use computational tools (many of which are open-source with less barrier to adoption) has spurred the analysis of these new data sources. Recent efforts include general frameworks for data wrangling (Wickham, 2014), improved access to high performance database systems (e.g., http://cran.rstudio.com/web/packages/dplyr/vignettes/databases.html), and sophisticated interfaces for data scraping and related web technologies (e.g., Nolan & Temple Lang, 2014). A number of teachers have taken heed of the advice of those working to incorporate data wrangling and management skills early in the curriculum (Carver & Stevens, 2014).

1.4.2 A Challenging Time for Statistics

Although this is undoubtedly an exciting time to be a statistician, there are a number of challenges that loom. The demand for quantitative skills is clearly there. The widely cited McKinsey report (Manyika et al., 2011) described the potential shortage of hundreds of thousands of workers with the skills to make sense of the enormous amount of information now available. But where will the hundreds of thousands of new workers anticipated by the McKinsey report and others come from? Graduates of undergraduate statistics programs will be a small fraction (even if the growth seen in the recent decade continues or accelerates). Increased supply is unlikely to be solved by an influx of new statistics doctoral students; while the number of doctoral graduates is slowly increasing, growth is insufficient to meet demand for new positions in industry, government, and academia.

Where else can these skilled graduates be found? If they aren't produced by statistics programs, where will they come from? The London report (Future of Statistical Sciences, 2013) describes the need for *data scientists*—the exact definition of which is elusive and very much a matter of debate—and raises important questions about the identity and role of statisticians (Horton, 2015; Wasserstein, 2015). What is meant by data scientist? What skills are required? What training is needed to be able to function in these new positions? What role does statistics have in this new arena? How can it be ensured that critical statistical messages be transmitted to students educated in other types of program that feed the data science shortfall?

A widely read Computing Research Association white paper (CRA, 2012) on the challenges and opportunities with "big data" starts in an encouraging manner: "The promise of data-driven decision-making is now being recognized broadly, and there is growing enthusiasm for the notion of 'Big Data'." But it is disconcerting that the first mention of statistics is on the sixth page of the report: "Methods for querying and mining Big Data are fundamentally different from traditional statistical analysis on small samples" (CRA, 2012, p. 6). The remaining references include statistics in passing as a bag of tricks (but not central to the use of data to inform decision-making).

In his video introduction to the keynote for the Strata + Hadoop Big Data Conference in 2015, United States President Barack Obama stated that "understanding and innovating with data has the potential to change how we do almost anything for the better." We applaud these sentiments. However, the fact that "statistics" was not mentioned in the presentation (or in many media depictions of this new and growing field) is a serious concern.

As another example of the challenges for the role of statistics in this era of "big data," consider the new Advanced Placement Computer Science Principles course, offered for the first time in the fall of 2016 (AP CS Principles, 2017). The response has been phenomenal, with almost 48,000 students taking the exam in 2017. In contrast, the AP Statistics exam was taken by about 7,000 students in its first year (1997) and about 216,000 students in 2017. This course focuses on the development of foundational computing skills, programming literacy, and an understanding of the impact of computing applications. It includes "creativity" and "data and information" as two of seven "big ideas" that underlie the curriculum. The description of creativity includes discussion of how "computing facilitates exploration and the creation of computational artifacts and new knowledge that help people solve personal, societal, and global problems" (AP CS Principles, 2017, p. 11). Big idea 3 is subtitled "Data and information facilitate the creation of knowledge." The course description states:

> Computing enables and empowers new methods of information processing, driving monumental change across many disciplines—from art to business to science. Managing and interpreting an overwhelming amount of raw data is part of the foundation of our information society and economy. People use computers and computation to translate, process, and visualize raw data and to create information. Computation and computer science facilitate and enable new understanding of data and information that contributes knowledge to the world. Students in this course work with data using a variety of computational tools and techniques to better understand the many ways in which data is transformed into information and knowledge. (p. 18)

Learning Objective 3.1.1 describes "use of computers to process information, find patterns, and test hypotheses about digitally processed information to gain insight and knowledge." This objective feels more expansive than the entire Advanced Placement Statistics course, where students are expected to "describe patterns and departures from patterns; plan and conduct a study; explore random phenomena using probability and simulation; and estimate population parameters and test hypotheses" (AP Statistics, 2016). While the success of its implementation remains to be seen, the Advanced Placement Computer Science Principles course provides an expansiveness of vision and pregnant sense of possibility for personal lives and the wider world. This is something that statistics education needs to learn from. The London report (Future of Statistical Sciences, 2013) warned that unless statisticians engage in related areas such as computation and data-related skills that are perhaps less familiar, there is a potential for the discipline to miss out on the important scientific developments of the twenty-first century.

What are the areas where statistics may need to adapt to be relevant to data science? In addition to pedagogy and content, technology is a key realm. While the Guidelines for Assessment and Instruction in Statistics Education (GAISE) K-12 (2005) and College (2016) reports encouraged the use of technology (which, on a more positive note, is now widespread in most courses), hundreds of thousands of high school students still use calculators rather than computers for their analyses, limiting their ability to move beyond simple calculations or gain any sense of realistic workflows that they might encounter in the real world. But much worse, it also narrowly constricts their vision of what statistics is and can be and neglects the huge potential of the visual sense for gaining insights from data. This is certainly not the technology being used by data scientists (or implemented in the new Advanced Placement Computer Science Principles course).

1.4.3 Where Are We Headed?

The growth of data science as a discipline presents both opportunities and challenges for statisticians and statistical educators (Ridgway, 2015). Data scientists are being hired by employers looking for innovative problem solvers with expertise in programming, statistical modeling, machine learning, and strong communication skills (Rodriguez, 2013).

Computer scientists bring useful skills and approaches to tackle the analysis of large, complex datasets. Statisticians bring important expertise in terms of the understanding of variability and bias to help ensure that conclusions are justified. In addition to “big data,” increasingly sophisticated probabilistic (stochastic) models are being developed, for example, in areas such as genetics, ecology, and climate science. Data science is often described as a “team sport.” The complementary skills from many historically disparate disciplines need to be blended and augmented to ensure that data science is on a solid footing. But this means that to be relevant in this age of data, statisticians must be better oriented toward data science, lest data science move on without statistics.

The emergence of statistics as a distinct discipline, and not just as an add-on to mathematics for highly educated specialists, is relatively new. The growth of data science has highlighted the importance of computer science and shifted the ground in terms of connections with other disciplines. Some aspects of statistics are rooted in mathematics. Moving forward, however, the connections to mathematics will remain rooted with aspects of discrete and applied mathematics along with the highly dynamic and productive interface with computer science is emphasized.

A number of individuals have proposed creative solutions for statisticians to respond to the data science challenge. In his 2012 ASA presidential address, Robert Rodriguez proposed a “big tent” for statistics that included anyone who uses statistics, including related disciplines such as analytics and data science (Rodriguez, 2013).

Brown and Kass (2009) warned that to remain vibrant, statistics needs to open up its view of statistical training. Nolan and Temple Lang (2010) outlined a curriculum to build computational skills as a basis for real-world analysis. Finzer proposed a framework to establish “data habits of mind” (2013). Diane Lambert of Google described the need for students to be able to “think with data” (Horton & Hardin, 2015).

The future of statistics is likely to be closely tied to aspects of data science. Success in this realm will require extensive and creative changes to our secondary and tertiary curriculum, along with partnerships with colleagues in related disciplines. Considerable work in the realm of statistics education research is needed to assess approaches and methods that attempt to address these capacities. The American Statistical Association has been proactive in creating reports to address potential curriculum changes for the present; see the GAISE College Report (2016) and GAISE K-12 Report (2005) and the Curriculum Guidelines for Undergraduate Programs in Statistical Science (ASA Undergraduate Guidelines Working Group, 2014). But statistics and data science are rapidly evolving, and curriculum and pedagogical changes need to evolve as well to remain relevant.

1.5 Closing Thoughts

There is a growing demand for statistical skills, knowledge, and competencies at the same time that the field of statistics and data science is broadening. Although there are many barriers to the adoption of changes in a curriculum that is already bulging with topics and increasingly heterogeneous in terms of approach, the alternative—allowing data science to proceed without statistics—is not attractive. It would not only diminish statistics, it would also diminish “data science” and worsen data-based decision-making in society. It would limit the attractiveness of statistics graduates to the employment market, and through that, limit the attractiveness of statistics programs themselves.

Cobb (2015) likened changing curricula to moving a graveyard: never easy in any circumstance. Developing additional capacities in statistics students takes time. This will likely require approaches that provide repeated exposure and a spiraling curriculum that introduces, extends, and then integrates statistical and data-related skills. It includes getting students to come to grips with multidimensional thinking, preparing them to grapple with real-world problems and complex data, and providing them with skills in computation. These are challenging topics to add to the curriculum, but such an approach would help students to tackle more sophisticated problems and facilitate their ability to effectively make decisions using data. Perhaps most importantly, it should also include successfully fostering an expansiveness of vision within students of the potential of statistics for their world and their own future lives.

With so much curricular, pedagogical and technological change under way, this is an exciting time to be involved in statistics education research. To chart our way into an exciting future of teaching and learning that best benefits our students, there are so many important research questions to be addressed, including determining how best to target, structure, teach, and assess the emerging curricula. There has never been a wider array of interesting and important problems for statistics education researchers to grapple with than there is right now. The insights within this volume should help spark and guide efforts in this realm for many years to come.

References

American Association for the Advancement of Science (2015). Meeting theme: Innovations, information, and imaging. Retrieved from https://www.aaas.org/AM2015/theme.

American Statistical Association Undergraduate Guidelines Workgroup. (2014). *Curriculum guidelines for undergraduate programs in statistical science*. Alexandria, VA: American Statistical Association. Online. Retrieved from http://www.amstat.org/asa/education/Curriculum-Guidelines-for-Undergraduate-Programs-in-Statistical-Science.aspx

AP Computer Science Principles. (2017). Course and exam description. Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap-computer-science-principles-course-and-exam-description.pdf.

AP Statistics. (2016). Course overview. Retrieved from https://apstudent.collegeboard.org/apcourse/ap-statistics/course-details.

Applebaum, B. (2015, May 21). Vague on your monthly spending? You're not alone. *New York Times*, A3.

Arnold, P. A. (2013). *Statistical Investigative Questions: An enquiry into posing and answering investigative questions from existing data*. Ph.D. thesis, Statistics University of Auckland. Retrieved from https://researchspace.auckland.ac.nz/bitstream/handle/2292/21305/whole.pdf?sequence=2.

Baldi, B., & Utts, J. (2015). What your future doctor should know about statistics: Must-include topics for introductory undergraduate biostatistics. *The American Statistician, 69*(3), 231–240.

Bartholomew, D. (1995). What is statistics? *Journal of the Royal Statistical Society, Series A: Statistics in Society, 158*, 1–20.

Box, G. E. P. (1990). Commentary. *Technometrics, 32*(3), 251–252.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science, 16*(3), 199–231.

Brown, E. N., & Kass, R. E. (2009). What is statistics? (with discussion). *The American Statistician, 63*(2), 105–123.

Carver, R. H., & Stevens, M. (2014). It is time to include data management in introductory statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the ninth international conference on teaching statistics*. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_C134_CARVER.pdf

Chambers, J. M. (1993). Greater or lesser statistics: A choice for future research. *Statistics and Computing, 3*(4), 182–184.

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education, 10*(3). Retrieved from http://www.amstat.org/publications/jse/v10n3/chance.html.

Cobb, G. W. (2015). Mere renovation is too little, too late: We need to rethink the undergraduate curriculum from the ground up. *The American Statistician, 69*(4), 266–282.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801–823.

Cohn, V., & Cope, L. (2011). *News and numbers: A writer's guide to statistics*. Hoboken, NJ: Wiley-Blackwell.

CRA. (2012). Challenges and opportunities with big data: A community white paper developed by leading researchers across the United States. Retrieved from http://cra.org/ccc/wp-content/uploads/sites/2/2015/05/bigdatawhitepaper.pdf.

De Veaux, R. D., & Velleman, P. (2008). Math is music; statistics is literature. *Amstat News, 375*, 54–60.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, England: Cambridge University Press.

Farrell, D., & Greig, F. (2015, May). *Weathering volatility: Big data on the financial ups and downs of U.S. individuals* (J.P. Morgan Chase & Co. Institute Technical Report). Retrieved from August 15, 2015, http://www.jpmorganchase.com/corporate/institute/research.htm.

Fienberg, S. E. (1992). A brief history of statistics in three and one-half chapters: A review essay. *Statistical Science, 7*(2), 208–225.

Fienberg, S. E. (2014). What is statistics? *Annual Review of Statistics and Its Applications, 1*, 1–9.

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education, 7*(2). Retrieved from http://escholarship.org/uc/item/7gv0q9dc.

Forbes, S. (2014). The coming of age of statistics education in New Zealand, and its influence internationally. *Journal of Statistics Education, 22*(2). Retrieved from http://www.amstat.org/publications/jse/v22n2/forbes.pdf.

Friedman, J. H. (2001). The role of statistics in the data revolution? *International Statistical Review, 69*(1), 5–10.

Friendly, M. (2008). The golden age of statistical graphics. *Statistical Science, 23*(4), 502–535.

Future of Statistical Sciences. (2013). *Statistics and Science: A report of the London Workshop on the Future of the Statistical Sciences*. Retrieved from http://bit.ly/londonreport.

GAISE College Report. (2016). *Guidelines for assessment and instruction in Statistics Education College Report*, American Statistical Association, Alexandria, VA. Retrieved from http://www.amstat.org/education/gaise.

GAISE K-12 Report. (2005). *Guidelines for assessment and instruction in Statistics Education K-12 Report*, American Statistical Association, Alexandria, VA. Retrieved from http://www.amstat.org/education/gaise.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2008). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8*(2), 53–96.

Grolemund, G., & Wickham, H. (2014). A cognitive interpretation of data analysis. *International Statistical Review, 82*(2), 184–204.

Hacking, I. (1990). *The taming of chance*. New York, NY: Cambridge University Press.

Hahn, G. J., & Doganaksoy, N. (2012). *A career in statistics: Beyond the numbers*. Hoboken, NJ: Wiley.

Hand, D. J. (2014). *The improbability principle: Why coincidences, miracles, and rare events happen every day*. New York, NY: Scientific American.

Holmes, P. (2003). 50 years of statistics teaching in English schools: Some milestones (with discussion). *Journal of the Royal Statistical Society, Series D (The Statistician), 52*(4), 439–474.

Horton, N. J. (2015). Challenges and opportunities for statistics and statistical education: Looking back, looking forward. *The American Statistician, 69*(2), 138–145.

Horton, N. J., & Hardin, J. (2015). Teaching the next generation of statistics students to "Think with Data": Special issue on statistics and the undergraduate curriculum. *The American Statistician, 69*(4), 258–265. Retrieved from http://amstat.tandfonline.com/doi/full/10.1080/00031305.2015.1094283

Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine, 2*, e124.

Kendall, M. G. (1960). Studies in the history of probability and statistics. Where shall the history of statistics begin? *Biometrika, 47*(3), 447–449.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Lawes, C. M., Vander Hoorn, S., Law, M. R., & Rodgers, A. (2004). High cholesterol. In M. Ezzati, A. D. Lopez, A. Rodgers, & C. J. L. Murray (Eds.), *Comparative quantification of health risks, global and regional burden of disease attributable to selected major risk factors* (Vol. 1, pp. 391–496). Geneva: World Health Organization.

Live Science. (2012, February 22). *Citrus fruits lower women's stroke risk*. Retrieved from http://www.livescience.com/18608-citrus-fruits-stroke-risk.html.

MacKay, R. J., & Oldford, R. W. (2000). Scientific method, statistical method and the speed of light. *Statistical Science, 15*(3), 254–278.

Madigan, D., & Gelman, A. (2009). Comment. *The American Statistician, 63*(2), 114–115.

Manyika, J., Chui, M., Brown B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation.

Marquardt, D. W. (1987). The importance of statisticians. *Journal of the American Statistical Association, 82*(397), 1–7.

Moore, D. S. (1998). Statistics among the Liberal Arts. *Journal of the American Statistical Association, 93*(444), 1253–1259.

Moore, D. S. (1999). Discussion: What shall we teach beginners? *International Statistical Review, 67*(3), 250–252.

Moore, D. S., & Notz, W. I. (2016). *Statistics: Concepts and controversies* (9th ed.). New York, NY: Macmillan Learning.

NBC News. (2011, January 4). *Walk faster and you just might live longer*. Retrieved from http://www.nbcnews.com/id/40914372/ns/health-fitness/t/walk-faster-you-just-might-live-longer/#.Vc-yHvlViko.

NBC News. (2012, May 16). *6 cups a day? Coffee lovers less likely to die, study finds*. Retrieved from http://vitals.nbcnews.com/_news/2012/05/16/11704493-6-cups-a-day-coffee-lovers-less-likely-to-die-study-finds?lite.

Nolan, D., & Perrett, J. (2016). Teaching and learning data visualization: Ideas and assignments. *The American Statistician* 70(3):260–269. Retrieved from http://arxiv.org/abs/1503.00781 .

Nolan, D., & Temple Lang, D. (2010). Computing in the statistics curricula. *The American Statistician, 64*(2), 97–107.

Nolan, D., & Temple Lang, D. (2014). *XML and web technologies for data sciences with R*. New York, NY: Springer.

Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature, 506*, 150–152. Retrieved from http://www.nature.com/news/scientific-method-statistical-errors-1.14700

Pfannkuch, M., Budget, S., Fewster, R., Fitch, M., Pattenwise, S., Wild, C., et al. (2016). Probability modeling and thinking: What can we learn from practice? *Statistics Education Research Journal, 15*(2), 11–37. Retrieved from http://iase-web.org/documents/SERJ/SERJ15(2)_Pfannkuch.pdf

Pfannkuch, M., & Wild, C. J. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 17–46). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Porter, T. M. (1986). *The rise of statistical thinking 1820–1900*. Princeton, NJ: Princeton University Press.

Pullinger, J. (2014). Statistics making an impact. *Journal of the Royal Statistical Society, A, 176*(4), 819–839.

Ridgway, J. (2015). Implications of the data revolution for statistics education. *International Statistical Review, 84*(3), 528–549. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/insr.12110/epdf

Rodriguez, R. N. (2013). The 2012 ASA Presidential Address: Building the big tent for statistics. *Journal of the American Statistical Association, 108*(501), 1–6.

Scheaffer, R. L. (2001). Statistics education: Perusing the past, embracing the present, and charting the future. *Newsletter for the Section on Statistical Education, 7*(1). Retrieved from https://www.amstat.org/sections/educ/newsletter/v7n1/Perusing.html.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.

Silver, N. (2014, August 25). *Is the polling industry in stasis or in crisis?* FiveThirtyEight Politics. Retrieved August 15, 2015, from http://fivethirtyeight.com/features/is-the-polling-industry-in-stasis-or-in-crisis.

Snee, R. (1990). Statistical thinking and its contribution to quality. *The American Statistician, 44*(2), 116–121.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.

Stigler, S. M. (2016). *The seven pillars of statistical wisdom*. Cambridge, MA: Harvard University Press.

Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician, 57*(2), 74–79.

Utts, J. (2010). Unintentional lies in the media: Don't blame journalists for what we don't teach. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics. Data and Context in Statistics Education*. Voorburg, The Netherlands: International Statistical Institute.

Utts, J. (2015a). *Seeing through statistics* (4th ed.). Stamford, CT: Cengage Learning.

Utts, J. (2015b). The many facets of statistics education: 175 years of common themes. *The American Statistician, 69*(2), 100–107.

Utts, J., & Heckard, R. (2015). *Mind on statistics* (5th ed.). Stamford, CT: Cengage Learning.

Vere-Jones, D. (1995). The coming of age of statistical education. *International Statistical Review, 63*(1), 3–23.

Wasserstein, R. (2015). Communicating the power and impact of our profession: A heads up for the next Executive Directors of the ASA. *The American Statistician, 69*(2), 96–99.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on *p*-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133.

Wickham, H. (2014). Tidy data. *Journal of Statistical Software, 59*(10). Retrieved from http://www.jstatsoft.org/v59/i10/.

Wild, C. J. (1994). On embracing the 'wider view' of statistics. *The American Statistician, 48*(2), 163–171.

Wild, C. J. (2015). Further, faster, wider. *The American Statistician*. Retrieved from http://nhorton.people.amherst.edu/mererenovation/18_Wild.PDF

Wild, C. J. (2017). Statistical literacy as the earth moves. *Statistics Education Research Journal, 16*(1), 31–37.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review, 67*(3), 223–265.

Chapter 2
What Is Statistics Education?

Andrew Zieffler, Joan Garfield, and Elizabeth Fry

Abstract Statistics education is an interdisciplinary field that is focused on the teaching and learning of statistics. This chapter describes how the discipline of statistics education has emerged and evolved from the training of statistics practitioners to the education of students at all levels and from a practice rooted in mathematics and science to a subject utilized across many disciplines. It also examines the current landscape of statistics education, exploring the diversity in the content and setting of statistics instruction around the world. Finally, the chapter outlines several opportunities and challenges on the horizon for statistics education.

Keywords Statistics education history • Secondary and tertiary levels • Statistics education reform • Professional organizations and journals • Content • Pedagogy • Technology • Research • Teacher preparation

2.1 Introduction

Statistics education is an interdisciplinary field that is focused on the teaching and learning of statistics. Evolving from the field of mathematics education, which supplied valuable theories of learning, models of conceptual development and change, and methods of qualitative research (e.g., teaching experiments, clinical interviews), statistics education has emerged as an independent area of inquiry and scholarship with its own journals, conferences, organizations, websites, and curriculum standards (Garfield & Ben-Zvi, 2008).

Perhaps 1982, the year of the First International Conference on Teaching Statistics (ICOTS), can serve as the official start date of this discipline. Prior to that time, very few people produced scholarship in statistics education, and the primary outlets for the dissemination of that scholarship were national and international conferences in mathematics education or mathematics education research journals. Scholars were also publishing statistics education research in other, domain-specific

A. Zieffler (✉) • J. Garfield • E. Fry
Department of Educational Psychology, The University of Minnesota, Minneapolis, MN, USA
e-mail: zief0002@umn.edu; jbg@umn.edu; fryxx069@umn.edu

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_2

journals (e.g., psychology, economics). Today, there is a growing community of scholars and researchers in statistics education, as well as an abundance of meetings, journals, and supporting organizations specifically associated with statistics education.

This chapter describes how the discipline of statistics education emerged and evolved from the training of statistics practitioners to the education of students at all levels, from a practice rooted in mathematics and science to a subject utilized across many disciplines. We examine the breadth of activities and resources in statistics education as well as current trends and future challenges. From an international perspective, a broad survey of statistical education will be provided, along with a synthesis of similarities across countries. We demonstrate how this discipline builds on and generates research, as well as how the research connects scholars and practitioners across many disciplines. Finally, we conclude with a discussion of current issues and challenges involving teachers, students, and researchers.

Before we begin, we note that although probability plays an important role in statistics education, we will rarely refer to it in this chapter. We made this decision in part because we view probability as a separate discipline from statistics and only a single component of statistics education, not its entirety. We refer readers more directly interested in the teaching of probability to Jones (2005), Kapadia and Borovcnik (1991), Shaughnessy (1991), Jones, Langrall, and Mooney (2007), and Chernoff and Sriraman (2014).

2.2 Brief History of Statistics Education

Although statistics education has evolved from the disciplines of mathematics education and science education, it nevertheless has its own history, which will be briefly described in this section. More descriptive and comprehensive accounts may be found in the writings of Bibby (1986a, 1986b), Hunter (1999), Neyman (1976), Scheaffer (2001), and Scheaffer and Jacobbe (2014).

2.2.1 Statistics Instruction: Late 1800s–Early 1900s

The teaching of statistics prior to 1900 was focused on the topics of collection, examination, and presentation of quantitative data (Bibby, 1986a; Fitzpatrick, 1955; Walker, 1890, 1929). The courses, taught at colleges and universities, were primarily intended to train government-sponsored researchers and professionals to enumerate and quantify characteristics of the populace. A typical statistics course of the era is described by Walker (1890):

> The pupil is taught to look up the data relating to a given subject, as these may be found scattered through a long series of official reports; to bring the various statements together; to examine them as to their proper compatibility; to test their accuracy by all means which may be available; and to put them back together into tables. The student is further taught to

> work out the percentages involved and to set one class of facts into relation with others; as, for example, to compute the ratio of valuation, or of expenditure, or of mortality, to each million or each thousand of the population concerned; and, finally, to make diagrams or charts, which shall exhibit graphically the several elements, taken in their due proportions, as ascertained by the investigation. (p. 7)

Many countries were also introducing statistics into their school curricula at early levels. For example, Bibby (1986a) suggests that in Hungary, probability was taught as part of the curriculum in schools as early as 1849 and in France was being taught in geography courses by 1868. In 1870, Britain established a Statistics in Schools committee to pursue the more formal introduction of statistics into the national curriculum (however, documents suggest it might have been disbanded after a mere 8 days). Japan and Belgium also introduced statistics into the school-level curriculum around this time period (Bibby, 1986b).

Around the turn of the twentieth century, the application of statistics trended to the natural sciences (e.g., biology) and began to emphasize increasingly specialized methodology rather than merely quantitative description. The content of statistics courses quickly followed suit. For example, many tertiary-level textbooks introduced the use and application of newer data analytic methods such as measures of center, dispersion, and correlation (e.g., Bailey, 1906; Davenport, 1899; King, 1912). At the secondary level, Perry (1900) proposed a more data-oriented syllabus for the mathematics curriculum being taught in British schools that included interpolation and probable errors.

It was around this time that statistics instruction became more mainstream in the curriculum at many institutions. Several colleges and universities began teaching courses solely devoted to statistics. In the next section, we explore some of the historical milestones of statistics instruction at the tertiary level.

2.2.2 Milestones in Tertiary-Level Statistics Instruction

In the early twentieth century, several universities (especially those in the United States) began teaching more formal statistics courses. These courses were taught in a diverse set of scholarly departments, most of which were employing statistics in the research and practical work of their respective academic fields. As the number of statistics courses being offered at the postsecondary level continued to rise, there were also more appeals for increased and better instruction from early proponents of statistics education (e.g., Chaddock, 1926; Willcox, 1910).

One of the most divisive arguments during this time period concerned the level and amount of mathematical theory that students needed in their statistical training. The growing division between courses offered in statistical application and in mathematical foundation played itself out in the public forum, with proponents for each side arguing the role of mathematics in statistics courses and the training of statisticians (e.g., Wilson, 1930). Hunter (1996) suggests that this is the beginning of a larger debate between two groups within the statistics community, "those beginning to explore the theoretical aspects of statistics … [and those] using statistics as a tool for work in other fields" (p. 14). This division would deepen over the next decade as statistics continued to free itself from other fields and become an independent discipline in its own right. The schism between theory and application (which ultimately birthed mathematical statistics; see David, 1998; Neyman, 1976; Stigler, 1996) was not unique to statistics and actually paralleled that taking place in mathematics (Craig, 1961).

The emphasis on mathematics was also taking place at universities outside the United States. For example, in a discussion at the Royal Statistical Society, Wishart points out that statistics was taught as part of the mathematical discipline at the University of Cambridge as early as 1931 and first awarded a diploma in mathematical statistics in 1948 (Pearson et al., 1955). In that article, Barnard and Bartlett reported similar emphasis within the statistics curriculum in the Universities of London and Manchester, respectively. In other countries, this shift didn't occur until later. China, for example, introduced mathematical statistics in the 1950s (Shi-Jian, 1990).

The progress and growth of mathematical statistics during the 1930s and 1940s shifted the introductory statistics curricula toward the inclusion of more advanced methods (e.g., Snedecor, 1948). For example, advancements in sampling theory (e.g., Neyman, 1934, 1938; Yates, 1946), as well as correlational and regression techniques (e.g., Bartlett, 1933; Fisher, 1924–1925; Tolley & Ezekiel, 1923; Welch, 1935), were all starting to be included in introductory courses. Additionally, by the early 1930s, Fisher developed analysis of variance as a practical method, and as Scheaffer (2001) so eloquently states, "(t)he influence of agriculture and Fisher on the maturing of statistics as a discipline cannot be overstated, and this influence permeates statistics education as well" (p. 2).

Rapid growth in the teaching, application, and use of statistical methods was coupled with a postwar influx of students in the undergraduate corridors (Bibby, 1986a; National Research Council, 1947). As veterans returned to civilian life, many chose to enroll in technical programs, including statistics. Although the numbers of statistics students were rising, by the end of the decade, the number of trained statisticians needed for industry positions far outpaced the applicant pool—which was sparse even before the war, causing many individuals and organizations to take notice (Balfour Committee, 1929; Cornell, 1945; Dwyer, 1945; Inman, 1990).

The proliferation of statistics courses and programs after World War II and a dearth in the number of trained and qualified users of statistics highlighted several challenges for statistics education, including the number of statistics programs that existed, the content that was/should be included in these programs, and how the programs were/should be organized (e.g., Hotelling, 1940; National Research Council, 1947; Pearson Committee, 1947). Questions about the preparedness of statistics teachers, as well as the growth in diversity and academic backgrounds of the students, were the focus of many of the publications of that period that addressed these challenges.

During this crisis, statistics education began to see a more organized effort from the broader statistical community, including the initiation of educational branches of the statistical societies and the formal development of undergraduate programs of statistics at many colleges and universities. For example, the Statistical Training Programme for Africa (STPA) was initiated in 1978 after an evaluation of training programs revealed that the existing programs could not keep pace with the need for trained statisticians (Tulya-Muhika, 1990).

The statistics curriculum at the tertiary level continued to evolve during the late part of the twentieth century. John Tukey's work on data analysis (Tukey, 1962, 1977) revolutionized both the practice and teaching of statistics. This work, coupled with increasing access to computers, was integral in moving the emphasis of many statistics courses from mathematical theory to data analysis, opening up the field to a broader population of students. More recent reform efforts of statistics instruction will be described in later sections.

2.2.3 Milestones in Secondary-Level Statistics Instruction

It took some time and effort to introduce topics of statistics into the secondary curriculum. Although probability was more connected to mathematics, the discipline in which statistics was eventually taught, these topics were not typically introduced until later in the twentieth century. From the turn of the century until the 1920s, the school-level mathematics curriculum in the United States, which was focused on algebra and geometry, was primarily geared toward preparing students for college. In 1929, the stock market crashed, and the resulting economic depression saw far fewer students attending college. As a result, schools shifted their educational focus to emphasize vocational and societal needs. During this time period, mathematics also became an elective subject, and subsequently schools worked to de-emphasize the sequential nature of the mathematics courses. This led to schools teaching new courses, a few of which began teaching statistics (Jones & Coxford, 1970).

The war effort of the mid-to-late 1940s brought a renewed emphasis to the study of mathematics at the school level. However, the content was aligned with the types of mathematics included on the induction test for the armed forces, again emphasizing algebra, geometry, and trigonometry. After the war, collegiate mathematicians expressed concern about the mathematical preparation of secondary students, and the College Entrance Education Board (CEEB) appointed a commission in 1954 to examine the mathematical needs of American students. One of the substantial recommendations made by the commission was to teach a course on probability during the last year of high school (Commission on Mathematics, 1959). Furthermore, the commission wrote a textbook for high school use, breaking with historic precedent of only making recommendations, and sanctioned a feasibility study of the curriculum.

As they also wrote the entrance examinations used for most colleges and universities, the CEEB's recommendations had a great deal of influence on secondary curriculum, and statistics instruction became much more prevalent at the secondary level. Their recommendations also affected school curricula in Canada, with many provinces rewriting their mathematics curriculum and including the study of probability and statistics (Crawford, 1970).

In the United Kingdom, virtually no probability or statistics was taught below the sixth form (ages 16–18; Green, 1982) prior to the 1960s. In 1959, the Organization for European Economic Co-operation organized a seminar on New Thinking in School Mathematics to take place at Royaumont, France. Eighteen countries participated with discussion focused on the reform of school-level mathematics (Shubring, 2013). The ideas and recommendations that emerged from this seminar ended up initiating an international pedagogical movement for mathematics. They also influenced future curricular content. The curricular recommendations issued in the resulting report, which influenced several broad educational projects, included the teaching of probability and statistical inference at the school level. Three projects that drew on this recommendation were the United Kingdom's School Mathematics Project, the Midlands Mathematics Experiment, and the Scottish Mathematics Group. All three curricular projects included statistical instruction, primarily the topic of probability, to students aged 11–16.

Throughout the 1960s, broad curricular projects addressing educational reform efforts continued to influence the mathematics taught at the secondary level. In 1967, the American Statistical Association (ASA) and the National Council of Teachers of Mathematics (NCTM) formed the Joint Committee on the Curriculum in Statistics and Probability to provide leadership and support for curricular efforts in grades K–12. Frederick Mosteller, a statistics professor at Harvard and then president of ASA, chaired the committee and was instrumental in many of its early successes. The first product that came directly out of the joint committee's efforts was *Statistics: A Guide to the Unknown* (Tanur et al., 1972), a book of solicited essays from statisticians positing the "value of statistics and probability and the contributions of these disciplines to the advancement of the biological, political, social, and physical sciences, as well as their usefulness in everyday life" (ASA, n.d.). The committee also produced a four-volume curricular series, *Statistics by Example* (Mosteller, Kruskal, Link, Pieters, & Rising, 1973), which included "real and interesting instructional material for teachers in high school to use in courses in statistics and probability" (American Statistical Association, n.d.). Both of these products further democratized statistics education by presenting statistical topics and analyses in an accessible manner to high school students and teachers.

John Tukey's (1977) book on exploratory data analysis had a large impact on the curriculum at the secondary level as well, where previously, only topics in probability had been introduced within the mathematics curriculum. One set of curricular materials influenced by Tukey's work was the Quantitative Literacy Project (QLP; see Scheaffer, 1990). QLP consisted of a series of books that were published in the late 1980s, again, as a joint project between the ASA and the NCTM. Around this time period, NCTM published a yearbook completely devoted to the teaching of statistics: *Teaching Statistics and Probability* (Shulte & Smart, 1981). A second NCTM yearbook related to teaching statistics was published in 2006 (Burrill & Elliott, 2006).

A sign that statistics was becoming a recognized strand of the mathematics curriculum was revealed in the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). These comprehensive and ambitious standards outlined goals for quality mathematics instruction in the United States. This document, which was influenced by a growing body of educational and psychological research that suggested that learning is an active, constructivist process and the changing role of technology in "doing" mathematics, called for increased instruction of statistics and probability at every grade level, K–12. The *Standards*, revised in 2000 (NCTM, 2000), have had a lasting impact on mathematics instruction for K–12 students and have also influenced curricular change at the tertiary level as well.

Prior to NCTM's success in publishing guidelines and standards detailing the role probability and statistics play in the mathematics curriculum, England successfully introduced statistics into their national secondary curriculum when the Schools Project in England published their report, *Mathematics Counts* (Cockcroft, 1982). This report was also famous for popularizing the term "numeracy." Scheaffer (2003) points out that this report may have been instrumental in helping countries such as Australia, New Zealand, and the United States write and adopt standards.

As secondary instruction and curriculum changed to meet these new standards, many statistics educators began to reconsider the introductory statistics course at the tertiary level in light of these shifts. The conversations and questions about how this course was taught, and the content that should be included, were again rekindled. In the next section, we examine milestones that affected more recent reform of statistics instruction at all educational levels.

2.2.4 Milestones in the Reform of Statistics Instruction

In the United States, the Mathematical Association of America organized a focus group to discuss and come up with recommendations for "reforming" the introductory college course. Cobb (1992) reported on the group's work, offering three recommendations to reform the teaching of introductory statistics: (1) emphasize statistical thinking, (2) include more data and concepts (less theory, fewer recipes), and (3) foster active learning. Building on these recommendations, Moore (1997) characterized reform needs in terms of changes in content (more data analysis, less probability), pedagogy (fewer lectures, more active learning), and technology (for data analysis and simulations).

Roughly 10 years after Cobb's report, the ASA funded a group of prominent statistics educators to write a set of instructional and assessment guidelines for teaching statistics at both the pre-K–12 and college levels. The resulting reports, the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE; ASA, 2005; Franklin et al., 2005), built on previous recommendations and reform efforts, as well as related curriculum standards to recent research on teaching and learning (see Franklin & Garfield, 2006). Both reports were endorsed by the ASA Board of Directors in 2005.

The pre-K–12 GAISE report focused on statistical literacy and laid out a statistical problem-solving framework across three levels based on students' development in statistical literacy. This framework included four components: (1) formulating a question, (2) collecting data, (3) analyzing data, and (4) interpreting results. It also encouraged teachers to focus on variation through all parts of the process. The college report provided six recommendations for the teaching and assessment of introductory statistics at the tertiary level: (1) emphasize statistical literacy and develop statistical thinking; (2) use real data; (3) stress conceptual understanding rather than mere knowledge of procedures; (4) foster active learning in the classroom; (5) use technology for developing conceptual understanding and analyzing data; and (6) integrate assessments that are aligned with course goals to improve as well as evaluate student learning. In 2016, the college report was updated to “reflect modern practice and take advantage of widely available technologies” (ASA, 2016, p. 6). In addition to the earlier six recommendations, this report emphasizes teaching statistics as an investigative process of problem-solving and decision-making and offering students experience with multivariable thinking.

Publication of the GAISE report inspired conversation and ideas. There have been sessions related to the GAISE report at major professional meetings such as the Joint Statistical Meetings, the Joint Mathematics Meetings, and the United States Conference on Teaching Statistics, every year since its publication. Additionally, many statistics teachers have developed courses that meet the recommended guidelines and have shared these implementations at professional meetings. In the United States, the National Science Foundation has funded new curriculum projects that build on the recommendations and learning outcomes presented in the GAISE report (e.g., Garfield, delMas, & Zieffler, 2012; Gould, Davis, Patel, & Esfandiari, 2010; Tintle et al., 2015; West, 2014; Woodard & McGowan, 2012).

Although many of the reform efforts described here were initiated in the United States, the impact of these efforts was not limited to the United States. One country that has worked to actively reform their school-level curriculum is New Zealand. This is in large part due to David Vere-Jones, who not only influenced statistics education in New Zealand but also at the international level. He was instrumental in New Zealand’s recognition of statistical learning outcomes (e.g., statistical literacy) within the school curriculum (see Vere-Jones, 1995).

More recently, in response to the widening gap between statistical practice and statistics education, a group of New Zealand researchers and statistics educators developed innovative computer-based approaches for teaching statistical inference (Pfannkuch, Forbes, Harraway, Budgett, & Wild, 2013). They found that the use of randomization and bootstrap approaches, along with dynamic visualizations, has the potential to make concepts of statistical inference more accessible to students at the secondary and tertiary levels. New Zealand has also been a leader in the preparation of statistics educators, preparing them to teach methods of data handling and data visualization, as well as simulation methods of inference (Forbes, Campos, & Helenius, 2013).

One of the most successful contributions to statistics education at the secondary level has been the United Kingdom’s Census at School project. This project engages grade 4–12 students with statistics by having them complete an online survey, analyze their class results, and compare them with results from other populations such as students in their own country and in other countries (American Statistical Association, 2017a). The Census at School project has also been adopted in several countries around the world (e.g., Australia, Canada, Ireland, Japan, New Zealand, and South Africa).

2.3 Professional Organizations and Journals Related to Statistics Education

Professional organizations/societies and journals within any discipline are essential for generating and sharing ideas and information. They help build energy within a professional community and provide recognition and support for members of the community. In this section, we describe some of the more prominent organizations/societies and journals dedicated to statistics education.

2.3.1 Organizations and Societies

Over time, the demand for statistics education and training has impacted many professional statistical organizations. Currently, almost all of the national statistical societies around the world explicitly include statistics education as part of their mission (e.g., Canada, India, Japan, New Zealand, the Philippines, South Africa). Some of these associations have initiated more formal committees or special interest groups dedicated to statistics education, which in turn have influenced the teaching of statistics. As an example, the New Zealand Statistical Association's statistics education committee has actively worked to include topics of data analysis at the primary and secondary school levels and to integrate the visionary work of statistics educators such as Chris Wild and Maxine Pfannkuch (see Wild, Pfannkuch, Regan, & Horton, 2011) into the school curriculum. They have also promoted the use of technology and graphical visualization in the teaching of data handling and problem-solving within the school curriculum.

In addition, many of these organizations sponsor projects, organize conference sessions, and offer participants networking opportunities specifically related to the teaching and learning of statistics. Three statistical organizations, the International Statistical Institute, the Royal Statistical Society, and the American Statistical Association, have long histories of supporting efforts in statistics education, and are highlighted below.

2.3.1.1 International Statistical Institute and the International Association for Statistical Education

The International Statistical Institute (ISI) formed a Committee on Statistical Education in 1948 to undertake educational activities in statistics and to collaborate with UNESCO and other UN agencies for this purpose (Vere-Jones, 1995). This committee was established on the advice of Dr. Stuart Rice, then President of the ISI, who was a strong advocate of statistics education. His 1949 paper, *Furtherance of Statistical Education* (Rice, 1949), would form the basis for the ISI's involvement in statistical education for the next several decades.

In the late 1970s, the ISI Committee on Statistical Education set up a number of task forces including the Task Force on Teaching Statistics at School Level

(TOTSAS) and the Task Force on International Conferences in Statistical Education. TOTSAS, initially chaired by Vic Barnett, established a regular newsletter available to interested school and university teachers of statistics. It also established the journal *Teaching Statistics*, which saw its first issue published in 1979. TOTSAS was also responsible for the publication of *Teaching Statistics in Schools Throughout the World* (Barnett, 1982), a survey of the "state of affairs" of statistics in preuniversity settings (e.g., where statistics was taught, example syllabi, etc.) around the world. It also published a companion volume, *The Training of Statisticians around the World* (Loynes, 1987).

The ISI Committee on Statistical Education became an official section of the ISI in 1991 and changed its name to the International Association for Statistical Education (IASE). Although its name changed, its focus remained that of improving the teaching and learning of statistics and advancing research in statistics education (Schuyten & Ottaviani, 2006). To carry out this mission, the IASE hosts a repeating 4-year cycle of professional meetings, (1) International Conference on Teaching Statistics (ICOTS), (2) Satellite and ISI Biennial, (3) Roundtable, and (4) Satellite and ISI Biennial. IASE also sponsors an International Statistical Literacy Project (ISLP) which includes international poster and project competitions.

2.3.1.2 Royal Statistical Society

In the United Kingdom, there is a long history of support for statistics education in the schools. One organization that has played a key role is the Royal Statistical Society (RSS). Initially interested in promoting the teaching of statistics at the college and university levels (e.g., RSS, 1947; Wishart, 1939), especially after World War II, the RSS was instrumental in including the teaching of statistics in the secondary school curriculum in England. Holmes (2003) gives a descriptive account of this history in his paper, *50 Years of Statistics Teaching in English Schools: Some Milestones*.

Pointing out the impact the RSS had on statistics education in the United Kingdom and maybe more broadly, Neville Davies (personal communication, March 20, 2016) writes:

> …the RSS was the first academic and professional statistical body to give sustained support for statistical education through budgeted funding (for its Centre) from 1995 to 2014. It gave overt and valuable backing for what it gradually came to believe to be a discipline in its own right. This support did, in fact, follow from the establishment of the first centre for statistical education based at Sheffield from 1982–1995, although the RSS had no involvement in that period at Sheffield.

The RSS continues to be involved with statistics education. In 2010 they acknowledged the coming of age of statistical education as a discipline when Chris Wild and his colleagues read a paper to the society (later published as Wild et al., 2011). This is noteworthy given that education-related papers are rarely published in RSS journals. The RSS also currently sponsors a campaign (getstats) which involves statistical literacy initiatives and resources to help teachers, journalists, and the general public to increase their statistical knowledge.

2.3.1.3 American Statistical Association

In the United States, the American Statistical Association (ASA) also increased its support of statistics education during the middle of the twentieth century. In 1947, ASA formally constituted and made the Committee on Training Statisticians the Section on Statistical Training. Although the choice of "training" rather than "education" was at the time deliberate, in 1972, Robert Hogg proposed changing the name of the section to the Section on Statistical Education. In 1973, when Hogg chaired the section, a new charter was submitted with the name change and approved by the section members that same year. See Inman (1990) for a lengthier history of the Section on Statistical Education. In addition to providing meeting space for statisticians interested in education, having a formal section also ensured that sessions related to statistical education would be included at the annual Joint Statistical Meetings.

Hogg continued to make statistics education a priority in the ASA. In 1991, he offered to be the program chair for the upcoming ASA winter meeting (Randles, 2007). The meeting, held in January of 1992 in Louisville, Kentucky, had a theme of statistics education. This meeting included over 600 participants (around 200 of whom were students) and resulted in collaborations that would impact statistics education for years to come (e.g., Cobb, 2013; Rossman & Garfield, 2011). Utts (2015) synthesizes many of the ASA's contributions to statistics education and also offers descriptions of their more recent contributions.

2.3.1.4 Professional Organizations in Mathematics

Several prominent professional organizations in mathematics also support statistics education. At the international level, the International Mathematical Union has long supported statistics education via its International Commission on Mathematical Instruction (ICMI). Statistics education researchers from around the world gather quadrennially at the International Congress on Mathematical Education (ICME). In 2008, the ICMI hosted a joint study with the IASE—Statistics Education in School Mathematics—Challenges for Teaching and Teacher Education (Batanero, Burrill, & Reading, 2011).

In the United States, statistics education has also been supported by the large and prestigious Mathematical Association of America (MAA). In addition to forming the focus group on the introductory course, chaired by George Cobb, which inspired the aforementioned GAISE reports, the MAA has published several books in their *Notes Series* on teaching statistics (e.g., Garfield, 2005; Moore, 2000), has sponsored workshops for mathematicians who teach statistics, and hosts a Special Interest Group in Statistics Education.

2.3.2 Journals

The presence of journals that publish scholarship in a particular field helps legitimize any discipline. There are currently four journals that are devoted primarily to statistics education. The oldest of these journals, *Teaching Statistics*, has been published since 1979 and consists of brief articles, activities, and research reports. Although based in the United Kingdom, the scope and reach of this journal are international, and its prestige has grown over the years as the journal changed in format, content, and scholarship.

The second oldest of the four journals, the *Journal of Statistics Education* (*JSE*), was established in 1993. Prior to this time, *Teaching Statistics* was the only domain-specific journal that published material related to statistics education. *JSE* was started to provide a peer-reviewed publication outlet for scholarship in statistics education and also to introduce innovation and curricular reform (Dietz, Gabrosek, Notz, & Short, 2013). In addition, *JSE* was the first electronic journal in statistics, having been published online since its inception.

The *Statistics Education Research Journal* (*SERJ*), a peer-reviewed, free electronic research journal of the IASE, made its debut in 2002, although its roots can be traced back to 1982 at the first International Conference on Teaching Statistics (ICOTS). At this meeting, the idea was kicked around to start a study group of researchers interested in statistics education, and the International Study Group for Research on Learning Probability and Statistics was born. In 1987, Joan Garfield took over as secretary of the study group from David Green—who had just become editor of *Teaching Statistics*—and wrote the group's first newsletter (Garfield, 1987). The newsletter was a catchall for information related to statistics education, including lists and, sometimes, descriptions of current published research, information about events and professional conferences, and other noteworthy tidbits. Efforts by two key figures in the international world of statistics education, Carmen Batanero and Maria Gabriella Ottaviani, convinced the IASE to include the study group as a special interest group and turn the newsletter into a regular publication. The newsletter officially changed names to the IASE *Statistical Education Research Newsletter* (*SERN*) in January of 2000. About this same time, a discussion began within the IASE about the problems, methodologies, and results that were stemming from statistics education research. This discussion, much of which was documented in *SERN* (e.g., Bacelar-Nicolau, 2001; Batanero, Garfield, & Ottaviani, 2001; Batanero, Garfield, Ottaviani, & Truran, 2000; Ottaviani, 2000), led to the establishment of *SERJ*. This journal is now published twice a year and continues to be sponsored by the IASE and ISI.

Technology Innovations in Statistics Education (*TISE*) is the newest publication to disseminate high-quality scholarship and research related to statistics education. It is an electronic journal founded by Rob Gould, which publishes research and other scholarship related to the use of technology to teach statistics at all educational levels, from kindergarten to graduate students and professionals. Unlike the other journals in statistics education, this journal is not affiliated with a professional organization, but, rather, is produced by the University of California, Los Angeles, whose statistics department has a strong commitment to improving education.

In addition to these four journals, there are also several journals (not specific to statistics education) that occasionally publish statistics education-related articles. For example, since its inception in 1947, *The American Statistician* has published many articles and special issues on topics such as the undergraduate curriculum, the teaching of Bayesian statistics in an introductory course, and the training of graduate teaching assistants. This journal also includes a regular section, called Teacher's Corner, comprised of peer-reviewed articles related to teaching statistics, primarily at the tertiary level.

The *International Statistical Review* has also published articles related to statistics education. These articles (some of which have been landmark articles in the field) have covered topics such as statistical thinking, assessment, literacy, and research on teaching and learning statistics (e.g., Garfield, 1995; Garfield & Ben-Zvi, 2007; Moore, 1997; Wild & Pfannkuch, 1999). International research journals in mathematics education such *as Mathematical Thinking and Learning*, *Mathematics Education Research Journal*, *Journal for Research in Mathematics Education*, and *Educational Studies of Mathematics* have published articles and special issues on research related to teaching and learning statistics. Some examples of these special issues include *The Role of Context in Developing Reasoning about Informal Statistical Inference* (Makar & Ben-Zvi, 2011) and *Statistical Reasoning: Learning to Reason from Samples* (Radford, 2015). In addition to the professional journals, some international handbooks produced by the mathematics education community have included comprehensive and influential syntheses of research that have contributed greatly to the shared knowledge base in statistics education (e.g., Langrall, Makar, Nilsson, & Shaughnessy, 2017; Shaughnessy, 1991, 2006).

It must be noted that, despite the prevalence of journals oriented specifically toward publishing scholarship on the teaching and learning of statistics, there are still questions about the legitimacy of these journals. For many faculty pursuing tenure and promotion in a department of mathematics or statistics, more weight is given to publication in discipline-specific journals, where scholarship related to theory is often emphasized over that related to teaching. Even for faculty in departments or programs geared toward mathematics education, publication in the statistics education journals may carry less weight.

2.4 The Current Landscape of Statistics Education

In many countries, statistics is taught, to some degree, at almost all educational levels, with most of the extended teaching of the subject occurring at the secondary and tertiary levels. At the secondary school level, the teaching of statistics generally takes place within the mathematics curricula, whereas at the tertiary level, statistics is taught across many disciplines and departments, including mathematics, engineering, psychology, sociology, public health, economics, and, of course, statistics. Over time, this diversity has led to questions about where statistics should be taught, the content that should be included, and who should be teaching it (e.g., Moore, 1988; Moore & Cobb, 2000).

Despite 20+ years of grappling with some of these questions, there is still abundant variation in the content and setting of statistics courses. This variation has become more apparent as the societal need for statistically literate citizens and a more statistically trained workforce increases and as we face an increasingly educationally diverse population of students in the classroom. To meet these challenges, statistics educators have primarily focused on updating the content and pedagogy in the classroom. The statistical content taught in schools is continually impacted by the changes in the practice and scope of the underlying discipline (see Chap. 1). For many years, the topics typically included in an introductory statistics course taught at the secondary or tertiary level included data collection, data representation, data summary, probability, and inference (Watson, 2006). This scope and sequence mirrored the basic process of analysis undertaken by practicing statisticians. Recently, advances in computing, which have had a major influence on statistical practice, have also led to changes in the classroom. For example, many instructors are using simulation methods to teach statistical inference and including more statistical modeling in their classrooms (e.g., Cobb, 2007; Garfield et al., 2012).

Statistics educators are also adopting pedagogical innovation such as activity-based learning, flipped classrooms, and collaborative learning that promotes student learning. These changes typically reflect the ideas that emerge from research on the teaching and learning of statistics (see Chap. 3). Many of these methods, ideas, and research are shared through the organizations, journals, and conferences described previously.

In the following sections, we explore the current landscape of statistics education at different educational levels. We first examine statistics education at the primary and secondary levels, focusing on the scope of content taught at these levels around the world, the standardized assessments that affect that content, and improving instruction via collaboration. We follow this up by examining statistics education at the tertiary level. There, we look at recommendations and reform around the undergraduate statistics major, and also at current thoughts about the introductory courses. Finally, we examine how statistics education is being utilized in the workplace.

2.4.1 Statistics Education at the School Level

Currently, there is a great deal of variation between countries in the role statistics plays in the primary and secondary curriculum. Some countries have nationally mandated curricula; others have a national curriculum on paper but no consistent implementation; and others have no national curriculum. One commonality across countries is that statistics content is frequently taught within the mathematics curriculum. For example, in Uganda, statistics is taught as part of applied mathematics (Opolot-Okurut & Eluk, 2011). In Ethiopia, statistics is taught as one of the five strands of mathematics (Michael & O'Connell, 2014), and in the Philippines, basic probability and statistics concepts are also taught within the mathematics curriculum (Reston & Jala, 2014). England also teaches statistics within their mathematics curriculum, but an inquiry by Smith (2004) recommended that statistics should be embedded in application subjects and taught by teachers of those subjects where it is applied. Ultimately, the government retained statistics instruction within the mathematics curriculum, a decision supported by the RSS (Porkess, 2011).

Recently there have been increased efforts worldwide to include statistical content at the primary, as well as, the secondary school level. For example, Australia set standards for learning statistics across all primary and secondary school years (Australian Curriculum, Assessment and Reporting Authority, 2013). Similarly, the New Zealand curriculum includes standards for teaching statistics as early as Year 1 of primary school (Ministry of Education, 2007). Ethiopia has also recently expanded statistics instruction from one unit taught in the grade 12 curriculum to include content throughout the K–12 experience (Michael & O'Connell, 2014). In contrast, school curriculum in the United States does not typically address statistical content until grade 6 (Watson, 2014), despite recommendations by the American Statistical Association to include statistical content earlier in the school curriculum (Franklin et al., 2005). The Common Core State Standards in Mathematics (Common Core State Standards Initiative, 2017) adopted by 42 out of 50 states as of May 2017 will, however, introduce ideas of data and measurement as early as grade 3.

As statistics is taught more widely throughout the school experience, professional organizations, governments, and universities have begun to collaborate to promote the teaching and learning of statistics. The Iranian Statistical Society, for example, along with the Isfahan Mathematics House, and the Mathematics Teachers' Society of Isfahan formed an annual team-based statistics competition for Iranian high school students. These organizations also collaborated on the development of a website in Farsi to promote the popularization of statistics (Parsian & Rejali, 2011). In the Philippines, these types of collaborations have led to the development of teaching materials—such as reference material for elementary school teachers to illustrate uses of statistics and secondary- and tertiary-level introductory statistics textbooks—as well as the implementation of a nationwide course to train teachers in probability and statistics (Reston & Bersales, 2011).

Not only is statistics included earlier in students' educational trajectories, but the scope of statistical content taught at the school level is arguably larger than it ever has been. For example, the middle school curriculum in France, which was once limited to a few methods of calculation and graphs, has expanded to include inferential topics such as sampling variability, probability, and simulation (Bihan-Poudec & Dutarte, 2014).

The availability of cheaper and more powerful technological tools has also broadened the scope of statistical content that is taught at the school level. For example, the prevalence of the graphing calculator in the 1990s made analysis (e.g., exploratory data analysis inference) more accessible for students, which in turn allowed more of this content to be included at the school level. More recently, access to more powerful computers has led New Zealand and the United States to use simulation methods to introduce statistical inference at the secondary grade levels (Forbes et al., 2013; Wild et al., 2011; Wild, Pfannkuch, Regan, & Parsonage, 2017).

Technology is not the only thing that influences curriculum at the school level. Curriculum is also influenced by a country's educational goals. One goal in many countries is the need to have a more statistically literate citizenry (e.g., Biggeri & Zuliani, 1999; Gal, 2004). Brazil, for example, emphasizes statistical reasoning as part of the civic formation of students (Campos, Cazorla, & Kataoka, 2011). In South Africa, statistical content is taught with the goal of preparing students for social and economic needs they will face as adults (Wessels, 2011).

Student performance on international assessments can also be a catalyst for broadening or streamlining school-level curriculum. Two such assessments, the Trends in International Mathematics and Science Study (TIMSS; Mullis & Martin, 2013) and the Programme for International Student Assessment (PISA; OECD, 2013), include a variety of statistical content questions (e.g., interpreting and representing data from graphs and charts, comparing characteristics of data sets, using data to make inferences) and have influenced several countries' curricular choices. Japan, for example, had removed most statistics content from their secondary-level curriculum in the early 2000s but, after their students performed poorly on the 2012 PISA, decided to re-expand their coverage of statistics (Fujii, Fukazawa, Takeuchi, & Watanabe, 2014). Poor student performance on PISA's statistical tasks also prompted Germany to emphasize data analysis and statistical reasoning at all grade levels of their national mathematics curricula (Martignon, 2011).

2.4.2 Statistics Education at the Tertiary Level

At the tertiary level, there are two primary sets of student stakeholders: (1) those pursuing a degree in statistics (major or minor) and (2) those who take statistics courses but are pursuing a degree in another field. In the United States, the number of students in both groups seems to be steadily increasing (Blair, Kirkman, & Maxwell, 2013; Bryce, 2002). Statistics educators have identified teaching and curricular challenges for both groups of students. Below, we attempt to describe both groups of students and some of the between-country variation in these students. We also identify some of the curricular challenges and considerations that arise in teaching statistics to these groups of students.

From 2010 to 2013, statistics was the fastest-growing undergraduate degree in science, technology, engineering, and mathematics (STEM) in the United States (ASA, 2015b), and this growth is not unique to the United States. However, whereas several countries have seen an increased interest in statistics as an undergraduate focus of study, there is still a great deal of between-country variation in the prevalence of the statistics major. For example, Richardson (2014) compared statistics majors at universities in Australia, Vietnam, and the United States and found that while many of the colleges and universities surveyed in Australia and the United States offer a statistics major, only four of the 50 colleges and universities surveyed in Vietnam offered a major in statistics. She also found variation in the content and requirements of statistics majors across countries. For example, she reports that in Vietnam about 45% of the courses for statistics majors are focused on mathematics and statistics, whereas in Australia and the United States, it is closer to 70%. This might be because most statistics majors in the United States and Australia are offered in mathematics departments.

Because of these differences, countries face several unique challenges. Recognizing the vast differences in the structure and content of the statistics major across institutions (even within a country), some professional statistical organizations have begun to consider how to provide more uniformity. In 1999, the American Statistical Association launched the Undergraduate Statistics Education Initiative (USEI) to focus organizational efforts, write guidelines, and provide marketing and continual support for undergraduate statistics programs within the United States. More recently, they also endorsed a set of curriculum guidelines for Undergraduate Programs in Statistical Science (ASA Undergraduate Guidelines Workgroup, 2014) that provides curricular suggestions about content and pedagogical considerations for colleges and universities that include a major, minor, or concentration in statistics. In Japan, where there were no statistics departments until 2014 (Kudo et al., 2014), the Japan Statistical Society recently wrote a set of certificate examinations for students and statisticians at many levels, from junior high to graduate level (Fujii et al., 2014).

While the number of students specializing in statistics at the tertiary level is producing more sophisticated data analysts—able to get a variety of jobs in business and industry—there is also more demand for students trained in other disciplines to understand basic ideas of data and chance and to be able to apply those concepts to their own area of study. The 2010 Conference Board of the Mathematical Sciences survey indicates that the enrollments in introductory statistics courses and upper-level statistics courses have both increased since 1995 (Blair et al., 2013). However, for many students pursuing a degree in a field other than statistics, especially those pursuing a non-STEM degree, the introductory course remains their only undergraduate exposure to statistics. As such, statistics educators have posed many questions about the goals, content, and instruction of this course. Below, we examine some of the goals, both cognitive and non-cognitive (e.g., attitudinal), that statistics educators have identified for this group of students.

Cognitive Goals. The GAISE guidelines (see Sect. 2.2.4) present one set of learning goals for students at the precollege levels and another set for those students at the tertiary level. In addition, statisticians involved in education have also provided their perspectives on what introductory statistics students should learn and understand about statistics. These are typically based on their own reflections about what they consider to be statistical concepts for educated citizens (e.g., Scheaffer, 2001; Utts, 2003, 2010).

Goals for introductory statistics students are often related to data and design. For example, these might include the design of an investigation, plan and collection of data, the exploration and comparison of observations, and the appropriate use of statistical inference (e.g., College Board, 2010). In addition, statistics educators have promoted goals of developing students' statistical literacy, statistical reasoning, and statistical thinking rather than rote skills, computations, and procedures (e.g., ASA, 2005; Batanero et al., 2011). Despite the fact that statistics educators commonly refer to statistical literacy, reasoning, and thinking, there is little agreement in statistics education scholarship about the operationalization of these outcomes. For example, Garfield and Ben-Zvi (2008) describe statistical literacy as being able to read and use basic statistical language and graphical representations to understand statistical information in the media and in daily life, whereas Gal (2002) argues that statistical literacy also encompasses the ability to critically evaluate and communicate statistical information and conclusions.

Non-cognitive Goals. Changing students' beliefs and attitudes about statistics, improving students' dispositions and motivation for studying statistics, and reducing students' anxiety about the subject have all been identified as goals in tertiary-level statistics courses (Wikipedia, 2016). Statistics education researchers have examined many of these non-cognitive outcomes for tertiary-level statistics students, often using specially designed instruments such as Student Attitudes Towards Statistics (Schau, Stevens, Dauphinee, & Del Vecchio, 1995).

Gal and Ginsburg (1994) recommend that statistics educators be aware of their students' reactions and feelings toward statistics, as they may have an effect on student learning. It is widely acknowledged that many students enter a statistics course with apprehension toward the subject, which works against their efforts to learn the material. Part of the problem is that many students experience mathematics anxiety and equate statistics with mathematics. In addition to overcoming anxiety and negative attitudes, certain dispositions (e.g., curiosity and awareness, skepticism) are needed in order to be successful in statistical work (Wild & Pfannkuch, 1999).

2.4.3 Statistics Education in the Workplace

Outside the classroom, the workplace is another common venue where statistical training and education take place. For example, many companies in the United States train their employees in quality control and improvement techniques (e.g., Six Sigma; Montgomery & Woodall, 2008). The workplace also offers a rich environment for statistics education researchers (e.g., Bakker, 2014; Bakker & Akkerman, 2015; Bakker, Kent, Derry, Noss, & Hoyles, 2008).

Many professions are also adopting more evidence-based practices and decision-making strategies; policy organizations, government, and healthcare are three examples. To better meet these needs, organizations like Statistics New Zealand (Janssen & Forbes, 2014) have developed a set of initiatives to enable them to make better decisions using data. Similarly, Awe and Vance (2014) describe Virginia Tech's Laboratory Interdisciplinary Statistical Analysis, which trains statisticians and scientists from third-world countries to communicate and collaborate with non-statisticians to make better evidence-based decisions.

2.5 Challenges and Opportunities in Statistics Education

Statistics education has changed a great deal since 1982, but there are still several issues and challenges that educators face in developing a statistically literate citizenry. Concerns about reproducibility of research (Peng, 2015; Wasserstein & Lazar, 2016), recent emphasis on data manipulation and computing (ASA Undergraduate Guidelines Workgroup, 2014), and abundance and variation in the types of data available for analysis (Gould, 2010) will all likely play a role in changing how statistics is taught. In this section, we outline several opportunities and challenges on the horizon for statistics education.

2.5.1 Technology

Technology has had, and will probably continue to have, the greatest impact on the teaching of statistics. Advances in technology have made the computation in statistics courses both more accessible and more accurate. This has reduced the amount of time statistics instructors need to spend on procedures, which allows for more time to be focused on statistical concepts (Chance, Ben-Zvi, Garfield, & Medina, 2008). For example, even the most basic statistical software makes the computation of confidence intervals and p-values trivial, and rather than devoting class time to reading and using statistical tables, such as the z and t tables (which are now obsolete), time instead can be devoted to students' understanding and interpretation of the values obtained. Technology tools are also being used to help students visualize and understand statistical concepts such as samples and p-values (e.g., via simulation; Biehler, Ben-Zvi, Bakker, & Makar, 2013).

More than just changing the method of computation or focus of learning within the statistics curriculum, technology has actually changed the curriculum itself. For example, entire statistics courses have been created around visualization methods, Monte Carlo simulation (Tintle et al., 2014; Zieffler & Catalysts for Change, 2013), and Bayesian methods (Albert & Rossman, 2009). These methods, which once were only possible with the computing power of a mainframe, can now even be taught at the introductory level.

Technology has also provided increasingly sophisticated tools to change the mode and environment in which statistics is taught. For example, many universities are exploring and expanding their capacities to teach statistics in an online environment. In addition to the universities, for-profit companies and institutions such as SAS, RStudio, and Statistics.com are also offering statistics courses and training online.

To meet the demand for statistics instruction, universities such as Johns Hopkins and Stanford have adopted online platforms (e.g., Coursera, iTunesU) to help facilitate course enrollments of several thousands of students. These platforms use technology to aid in course administration and student feedback to make these large enrollments possible. For example, Carnegie Mellon's Open Learning Initiative features an online virtual tutor that gives immediate and targeted feedback to students (Lovett, Meyer, & Thille, 2008).

It is important to note that technology has also allowed for greater access to teaching materials and scholarship related to statistics education. For example, CAUSEweb, the Consortium for the Advancement of Undergraduate Statistics Education's website, is a comprehensive web repository that provides a multitude of peer-reviewed resources and professional development opportunities (such as webinars) for statistics instructors (Consortium for the Advancement of Undergraduate Statistics Education, 2017). In addition, CAUSE sponsors the biannual United States Conference on Teaching Statistics, as well as two electronic conferences: The Electronic Conference on Teaching Statistics and the Electronic Undergraduate Statistics Research Conference. The IASE has posted many valuable links and resources on their website, including past conference and roundtable proceedings, recent dissertations in statistics education, and relevant conferences and meetings, as well as an extensive website of resources for the previously mentioned International Statistical Literacy Project (IASE, 2017a, 2017b). The ASA Center for Statistics Education also hosts several websites with resources for teachers including many activities and lesson plans (e.g., ASA, 2017b, 2017c).

2.5.2 Pedagogy

Recommendations for the teaching and assessment of introductory statistics have been clear that instructors need to adopt a less lecture-dominated classroom when teaching statistics (ASA, 2016; Saxe & Braddy, 2015). Student-centered pedagogical methods such as active learning, flipped classrooms, and project-based learning are potential alternatives to the traditional lecture format. At the elementary and secondary levels, some educators have used the perspective of learning environments to provide a holistic integration to the teaching of statistics (see Chap. 16).

Active learning methods are an alternative to a lecture-based classroom. The use of active learning or student discovery in teaching is not unique to the statistics classroom. And, in general, recommendations for teaching statistics draw heavily from evidence accumulated from the broader fields of psychological and educational research. Several examples of using active learning for teaching statistics are readily available, including recent curricular efforts by Chance, Wong, and Tintle (2016), Garfield et al. (2012), and Rossman and Chance (2013).

Another student-centered approach to instruction is the use of a flipped classroom. In a flipped classroom, students read content or watch videos outside of class and then spend class time working with other students on the material. There has been some evidence that tertiary-level students in introductory statistics courses that use the flipped classroom approach have had better performance and attitudes than students in comparable lecture-based courses (e.g., Wilson, 2013; Winquist & Carlson, 2014).

The use of student projects is another pedagogical method for creating a student-centered classroom. Not only can projects help students see statistics as a valuable part of the investigative process of problem-solving and decision-making (ASA, 2016), but they can also provide instructors with a more authentic manner in which to assess students' broader understanding of statistics (e.g., Fillebrown, 1994; Zeleke, Lee, & Daniels, 2006).

As awareness of these instructional recommendations spreads, it is important to examine how responsive statistics instructors are to calls for change. Ten years after the "Cobb Report" was issued, Garfield, Hogg, Schau, and Whittinghill (2002) surveyed statistics instructors to determine the extent to which reform efforts were impacting the introductory course. The results suggested that very little had changed in the teaching of statistics outside of an increased use of technology. More recently, a survey of introductory statistics instructors at the tertiary level revealed that many instructors are engaging in recommended practices, such as having students use technology, assessing conceptual knowledge, and using real data. However, the results also indicated that most instructors continue to use lecture as the predominant method of content delivery, not using recommended pedagogical practices such as use of collaborative activities that encourage students to experience and construct concepts (Fry & Garfield, 2015; Garfield, delMas, Zieffler, & Fry, 2015). When asked why, many instructors reported hesitation or resistance to make changes in content delivery because of constraints such as personal time limitations and student characteristics.

2.5.3 More Statistics at the School Levels

We continue to see more and more statistical contents trickle down to the school-level curriculum. In the United States, the Common Core State Standards include more statistics content than most schools have previously taught. Countries like Australia, Ethiopia, Israel, and New Zealand are also increasing school-level exposure to statistics content. This expansion of the statistics curricula is also coupled with the success of programs such as the Advanced Placement (AP) program, in the United States, which offers tertiary-level curricula to secondary school students. As part of the AP program, students can opt to take an examination, which if they score well on, can earn them course credit at some colleges and universities. This program, which started in 1997, has had prolific success, administering 7500 exams in its inaugural year and 185,000 exams in 2014 (Rossman, 2015, personal communication), a roughly 25-fold increase! Histories of the development of AP Statistics are offered in Roberts, Scheaffer, and Watkins (1999) and Franklin et al. (2011).

As more statistics is taught at the school level, there are many questions about the amount and type of statistical content that should be taught at these levels. For example, many of the more recent Statistical Reasoning, Thinking, and Literacy (SRTL) international research forums have focused on informal statistical inference and statistical modeling (SRTL, 2017), and several participants were engaged in the study of students' and teachers' reasoning about these ideas at the school level (e.g., see Zieffler & Fry, 2015). There are also open questions about how the tertiary content will have to change to accommodate students that have already learned most of what is currently covered in an introductory postsecondary course in statistics. Finally, and perhaps most importantly, there are questions about the preparedness of teachers at the primary and secondary levels to teach this content. It is there that we turn next.

2.5.4 Teacher Preparation and Development

Issues of how to prepare and support teachers at all educational levels continue to be of major importance in statistics education (see, e.g., Batanero et al., 2011). At the primary and secondary levels, teachers, many of whom have never studied statistics, need to be prepared to teach the expanded statistical content now included in the curriculum. Even mathematics teachers, who generally have some statistics training in their background, will need additional preparation and professional development beyond their mathematical training. This might include development of their statistical content knowledge and instruction on the use of appropriate tools and technology, as well as on educationally recommended pedagogical methods for teaching statistics (e.g., active learning, use of student projects, etc.) (see also Chaps. 12 and 13).

In some countries, collaborations exist among different entities (e.g., government institutions, professional statistics associations, academic institutions, private organizations) to train teachers to teach statistics. For example, the Iranian Statistical Society, along with the Iranian Association of Mathematics Teachers' Societies, convinced the Iranian Ministry of Education to add a statistics course in the national high school curriculum in order to promote statistical literacy. Professional organizations along with the ISS organized lectures and workshops throughout the country to train mathematics teachers to teach statistics and provided various resources (Parsian & Rejali, 2011). In the Philippines, individuals, universities, the government, and private organizations have all worked together to prepare teachers to implement a revised school curriculum, which includes statistics and probability (Reston & Bersales, 2011). Reform efforts have also included the reviews of locally written textbooks, grants for writing reference books, and forums for presenting research in statistics.

Universities in different countries have also responded to the need to train future statistics teachers. For instance, Froelich (2011) describes a new curriculum in statistical content for future secondary mathematics teachers at a large university in the United States. The curriculum engages future teachers with data collection and analysis, probability, and inferential statistics and emphasizes similarities and differences between mathematics and statistics. Similarly, Green and Blankenship (2013) developed an introductory statistics course for preservice elementary teachers to help emphasize the importance of statistics in the elementary curriculum. In Germany, six Universities of Education have collaborated to have future teachers participate in regular seminars on both statistical and pedagogical content knowledge (Martignon, 2011). In addition, some German states include statistical questions on data analysis and visualization in their teaching certification examinations.

One challenge in teacher development, to date, has been the lack of guidance about the statistical content that teachers should be prepared to teach. Although previous reports, such as the *Mathematical Education of Teachers Report* (Conference Board of the Mathematical Sciences, 2012), documented the importance of statistical training for mathematics teachers, there was no specific guidance about content. In April of 2015, the ASA published the *Statistical Education of Teachers* (*SET; American Statistical Association, 2015a*), in which the statistical training of K–12 teachers was more defined. This document emphasized the need for teachers to understand content that extends beyond the level they teach. The *SET* document outlines specific topics and concepts for teachers at each of these levels and advocates for at least one statistics course for teachers that integrates both content and pedagogical knowledge.

There are also several other challenges that have been documented related to teacher development. For statistics teachers in many countries, their undergraduate coursework has not prepared nor supported them in becoming effective teachers of statistics. Whether this is due to outdated content knowledge (Reston & Jala, 2014), lack of pedagogical training (Sorto, 2011), or economic conditions that constrain the available resources to train teachers (Muñoz, Arañeda, Sorto, & León, 2014), these challenges pose real problems for many countries where statistics has become a focus of the curriculum in both primary and secondary schools.

In response to these challenges, many countries have implemented professional development opportunities for their in-service and preservice teachers. For example, Reston and Jala (2014) report on the use of workshops in the Philippines that promote reflective practice intended to help improve the teaching practices of statistics teachers. Similarly, Muñoz et al. (2014) write about the curricular materials that have been developed to expose Chile's future elementary teachers to more statistical content and improve their knowledge related to statistics. Although professional development for preservice and in-service teachers appears to offer a short-term method of improving the teaching of statistics, more preparation and study are needed to develop statistics educators who can lead the field forward.

2.5.5 Graduate-Level Study in Statistics Education

Although graduate students have, for some time, been completing doctoral research that focuses on the teaching and learning of statistics (many dissertations are archived on the IASE website; IASE, 2017b), it is only in the last decade that universities have developed graduate-level courses, research seminars, and specialized research mentoring in statistics education. Certification programs are another way in which educators engage in graduate-level training. The RSS offered a Certificate in Teaching Statistics in Higher Education which was a unique postgraduate qualification to provide personal development for teachers of statistics in tertiary-level institutions.

The University of Minnesota initiated the first formal graduate program in statistics education in 2002. Prior to that, there were no institutions that offered a master's degree and/or Ph.D. in statistics education. Since that time, other institutions in the United States have also developed a graduate program in statistics education (e.g., University of Florida, Portland State University, University of Georgia). Depending on the institution, these programs are located either in statistics departments or in education departments. In addition, graduate students at a variety of universities around the world continue to complete dissertations involving statistics education research (e.g., Open University, The Netherlands; Pontificia Universidade de Sao Paulo, Brazil; University of Granada, Spain; The University of Haifa, Israel). As requirements for doctoral degrees vary from country to country, most programs where students write a dissertation in statistics do not actually include coursework or seminars in this area.

In 2009 the ASA approved a set of guidelines for graduate programs in statistics education (ASA, 2009). It is noteworthy that these guidelines include the need for expertise in the content of statistics, the practice of statistics, the teaching of statistics, and the methods of conducting educational research. These guidelines also recommended that graduate committees for students conducting research on the teaching and learning of statistics include faculty from both education and statistics departments.

2.6 Conclusion

It is exciting to see statistics embraced by the public and popularized in the media, in part due to people such as Hans Rosling and Nate Silver. Rosling's lectures for Technology, Entertainment, Design (TED) feature colorful graphics and dynamic data representation to tell stories of complex multidimensional data. Rosling is also one of the founders of Gapminder (2017), a nonprofit venture to increase people's use and understanding of statistics and other information about social, economic, and environmental development throughout the world. American writer Nate Silver gained visibility for himself and for statistics when he accurately predicted the outcome of the 2008 US presidential election in 49 out of 50 states. Silver is currently the editor in chief of ESPN's FiveThirtyEight blog, which publishes articles analyzing statistics information in politics, economics, and sports (ESPN, 2017).

Statistics also gained visibility worldwide in 2013, during the International Year of Statistics. This initiative was sponsored by hundreds of organizations and institutes around the world, including the International Statistical Institute, which was also one of the initiators of this handbook. The International Year of Statistics had the goals of raising public awareness, introducing young people to statistics careers, and promoting creativity in statistical science. As part of this initiative, a website was launched with information to educate the public about the discipline of statistics, information about careers in statistical sciences, and resources for teachers of statistics (ASA, 2017d).

As the proliferation of data increases and the importance of data-based decisions becomes more widely recognized, there will most likely be even more of a demand for statistically literate citizens. The challenges to the statistics education community to stay current will also increase due to the rapidly changing discipline of statistics as well as the advent of several new data science programs and degrees (Swanstrom, 2017).

It is incumbent on statistics educators, now more than ever, to continue to ensure that students are being taught effectively, with an emphasis on learning. Technology will continue to transform the statistics curriculum and statistical learning environment, and statistics educators must be willing to adopt promising pedagogical and curricular innovation. This means that researchers need to continue to study these innovations, paying particular attention to the mechanisms and supports underlying successful classroom implementation. Additionally, professional organizations and conference organizers need to continue to support educational efforts and teacher development, not only through funding but also by making resources and materials available and accessible.

We also hope that the trend toward student-centered classrooms and pedagogy extends to statistics courses beyond the introductory course. As data science and statistics programs develop and expand their student base, it will be critical that we not only evaluate our teaching efforts in these advanced courses but also that we are studying the integration of these courses with the introductory course. Data science has the promise of attracting students from many different domains and backgrounds. As such, we also need to monitor the education pipeline, to work toward inclusion of students rather than exclusion.

The knowledge and experience accumulated to date by the statistics education community, along with published priorities and guidelines, have set statistics education up for a bright future. With institutions developing new programs and courses, the field should continue to grow and improve, promoting more and higher-quality research and providing a solid, recognized foundation for the discipline of statistics education.

References

Albert, J., & Rossman, A. (2009). *Workshop statistics: Discovery with data, a Bayesian approach*. Emeryville, CA: Key College Press.

American Statistical Association. (2005). *Guidelines for assessment and instruction in statistics education: College report*. Alexandria, VA: Author. Retrieved from http://www.amstat.org/education/gaise/

American Statistical Association. (2009). *Statistics education graduate programs*. Alexandria, VA: Author.

American Statistical Association. (2015a). *The statistical education of teachers*. Alexandria, VA: Author. Retrieved from http://www.amstat.org/education/SET/SET.pdf

American Statistical Association. (2015b). Statistics: Fastest-growing undergraduate STEM degree. *AmStat News*. Alexandria, VA: Author. Retrieved from http://magazine.amstat.org/blog/2015/03/01/statistics-fastest-growing-undergraduate-stem-degree.

American Statistical Association. (2016). *Guidelines for assessment and instruction in statistics education: College report*. Alexandria, VA: Author. Retrieved from http://www.amstat.org/education/gaise/

American Statistical Association. (2017a, May 23). *Census at school—United States*. Retrieved from http://ww2.amstat.org/censusatschool/about.cfm.

American Statistical Association. (2017b, May 23). *STatistics Education Web (STEW)*. Retrieved from http://www.amstat.org/ASA/Education/STEW/home.aspx.

American Statistical Association. (2017c, May 23). *Welcome to stats 101: A resource for teaching introductory statistics*. Retrieved from http://community.amstat.org/stats101/home.

American Statistical Association. (2017d, May 23). *World of statistics*. Retrieved from http://www.worldofstatistics.org/.

American Statistical Association. (n.d.). *Frederick Mosteller: 1916–2006*. Statisticians in History [website]. Retrieved from http://www.amstat.org/about/statisticiansinhistory/index.cfm?fuseaction=main.

American Statistical Association Undergraduate Guidelines Workgroup. (2014). *Curriculum guidelines for undergraduate programs in statistical science*. Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/education/curriculumguidelines.cfm

Australian Curriculum, Assessment and Reporting Authority. (2013). *The Australian curriculum, Version 5.0, 20 May 2013*. Sydney, NSW: Author.

Awe, O., & Vance, E. (2014). Statistics education, collaborative research, and LISA 2020: A view from Nigeria. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_5A3_AWE.pdf

Bacelar-Nicolau, H. (2001). On the paper 'Research in statistical education: Some priority questions'. *Statistical Education Research Newsletter, 2*(1), 3–4. Retrieved from http://iase-web.org/documents/SERJ/Newsjan01.pdf

Bailey, W. B. (1906). *Modern social conditions: A statistical study of birth, marriage, divorce, death, disease, suicide, immigration, etc., with special reference to the United States*. New York: The Century.

Bakker, A. (2014). Characterising and developing vocational mathematical knowledge. *Educational Studies in Mathematics, 86*(2), 151–156.

Bakker, A., & Akkerman, S. F. (2015). A boundary-crossing approach to support students' integration of statistical and work-related knowledge. *Educational Studies in Mathematics, 86*(2), 223–237.

Bakker, A., Kent, P., Derry, J., Noss, R., & Hoyles, C. (2008). Statistical inference at work: Statistical process control as an example. *Statistics Education Research Journal, 7*(2), 130–145. Retrieved from https://iase-web.org/documents/SERJ/SERJ7(2)_Bakker.pdf

Balfour Committee. (1929). *Report of the committee on industry and trade*. London: Her Majesty's Stationery Office.

Barnett, V. (Ed.). (1982). *Teaching statistics in schools throughout the world*. Voorburg, The Netherlands: International Statistical Institute.

Bartlett, M. S. (1933). On the theory of statistical regression. *Proceedings of the Royal Society of Edinburgh, 53*, 260–283.

Batanero, C., Burrill, G. F., & Reading, C. (Eds.). (2011). *Teaching statistics in school mathematics—Challenges for teaching and teacher education: A Joint ICMI/IASE Study: The 18th ICMI Study*. Dordrecht, The Netherlands: Springer.

Batanero, C., Garfield, J. B., & Ottaviani, M. G. (2001). Building a research agenda for statistics education. A response to reactions in SERN 2(1). *Statistical Education Research Newsletter, 2*(2). Retrieved from http://iase-web.org/documents/SERJ/Newsmay01.pdf.

Batanero, C., Garfield, J. B., Ottaviani, M. G., & Truran, J. (2000). Research in statistical education: Some priority questions. Statistical Education Research Newsletter, 1(2), 2–6. Retrieved from http://www.ugr.es/~batanero/sergroup.htm.

Bibby, J. (1986a). *Notes towards a history of teaching statistics*. Edinburgh, Scotland: John Bibby (Books).

Bibby, J. (1986b). 1786–1986: Two centuries of teaching statistics. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 478–493). Victoria, BC: The Organizing Committee, ICOTS2.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–690). New York: Springer.

Biggeri, L., & Zuliani, A. (1999). *The dissemination of statistical literacy among citizens and public administration directors*. Paper presented at the 52nd Session of the International Statistical Institute, Helsinki, Finland. Retrieved from http://iase-web.org/documents/papers/isi52/bigg0981.pdf.

Bihan-Poudec, A., & Dutarte, P. (2014). What did they learn? Statistics skills: From French secondary school to university. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_2D1_BIHANPOUDEC.pdf

Blair, R., Kirkman, E., & Maxwell, J. (2013). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2010 CBMS Survey*. Washington, DC: American Mathematical Society.

Bryce, G. R. (2002). Undergraduate statistics education: An introduction and review of selected literature. *Journal of Statistics Education, 10*(2). Retrieved from http://www.amstat.org/publications/jse/v10n2/bryce.html.

Burrill, G., & Elliott, P. C. (Eds.). (2006). *Thinking and reasoning with data and chance: 68th NCTM Yearbook 2006*. Reston, VA: National Council of Teachers of Mathematics.

Campos, T. M., Cazorla, I. M., & Kataoka, V. Y. (2011). Statistics school curricula in Brazil. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 9–14). New York: Springer.

Chaddock, R. (1926). The function of statistics in undergraduate training. *Journal of the American Statistical Association, 21*(153), 1–8. Retrieved from https://www.asapresidentialpapers.info/ASA_1925.htm

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2008). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1). Retrieved from http://escholarship.org/uc/item/8sd2t4rr.

Chance, B., Wong, J., & Tintle, N. (2016). Student performance in curricula centered on simulation based inference: A preliminary report. *Journal of Statistics Education, 24*(3), 114–126.

Chernoff, E. J., & Sriraman, B. (Eds.). (2014). *Probabilistic thinking: Presenting plural perspectives*. New York: Springer.

Cobb, G. W. (1992). Teaching Statistics. In L. Steen (Ed.), Heeding the call for change: Suggestions for curricular action (pp. 3-43). Washington: Mathematical Association of America

Cobb, George W. (2007). The introductory statistics course: A Ptolemaic curriculum?. Technology Innovations in Statistics Education, 1(1). Retrieved from: http://escholarship.org/uc/item/6hb3k0nz

Cobb, G. W. (2013). What might a twenty-year old conference tell us about the future of our profession? Journal of Statistics Education, 21(2). http://ww2.amstat.org/publications/jse/v21n2/cobb.pdf

Cockcroft, W. (1982), Mathematics counts: Report of the Committee of Inquiry into the teaching of mathematics in schools. London: HMSO. http://www.educationengland.org.uk/documents/cockcroft/

College Board. (2010). AP Statistics course description. Princeton, NJ: Author.

Commission on Mathematics. (1959). Program for college preparatory mathematics. New York: College Entrance Examination Board.
Common Core State Standards Initiative. (2017, May 23). *Mathematics standards*. Retrieved from http://www.corestandards.org/Math/.
Conference Board of the Mathematical Sciences. (2012). *The mathematical education of teachers II*. Providence, RI and Washington, DC: American Mathematical Society and Mathematical Association of America.
Consortium for the Advancement of Undergraduate Statistics Education. (2017, May 23). *CAUSE*. Retrieved from https://www.causeweb.org/.
Cornell, F. G. (1945). Training to supply the demand. *Journal of the American Statistical Association, 40*(230), 167–171.
Craig, C. C. (1961). The first course in mathematical statistics. *The American Statistician, 15*(5), 14–16. 25.
Crawford, D. H. (1970). Rethinking school mathematics: 1959–present. In P. S. Jones (Ed.), *A history of mathematics education in the United States and Canada, 32nd Yearbook of the National Council of Teachers of Mathematics* (pp. 426–450). Washington, DC: National Council of Teachers of Mathematics.
Davenport, C. B. (1899). *Statistical methods with special reference to biological variation*. London: Chapman and Hall.
David, H. A. (1998). Statistics in U.S. universities in 1933 and the establishment of the Statistical Laboratory at Iowa State. *Statistical Science, 13*(1), 66–74.
Dietz, E. J., Gabrosek, J., Notz, W., & Short, T. H. (2013). Celebrating 20 years of the Journal of Statistics Education. *Journal of Statistics Education, 21*(2). Retrieved from http://www.amstat.org/publications/jse/v21n2/editor_reflections.pdf.
Dwyer, P. S. (1945). Report on the Rutgers meeting. *The Annals of Mathematical Statistics, 16*(3), 317–318.
ESPN. (2017, May 23). *FiveThirtyEight*. Retrieved from https://fivethirtyeight.com/.
Fillebrown, S. (1994). Using projects in an elementary statistics course for non-science majors. *Journal of Statistics Education, 2*(2), 1–5.
Fisher, R. A. (1924–1925). *Note on Dr. Campbell's alternative to the method of least squares*. Unpublished manuscript, Barr Smith Library, University of Adelaide, Australia.
Fitzpatrick, P. J. (1955). The early teaching of statistics in American Colleges and Universities. *The American Statistician, 9*(5), 12–18.
Forbes, S., Campos, P., & Helenius, R. (2013). Promoting statistics to youth through the International Statistical Literacy Project. In *Proceedings of the 59th World Statistics Congress of the International Statistical Institute, 2013*. The Hague, The Netherlands: International Statistical Institute. Retrieved from http://www.statistics.gov.hk/wsc/IPS068-P1-S.pdf
Franklin, C., & Garfield, J. B. (2006). The GAISE Project: Developing statistics education guidelines for pre K-12 and college courses. In G. Burrill (Ed.), *Thinking and reasoning with data and chance: 2006 NCTM yearbook* (pp. 345–375). Reston, VA: National Council of Teachers of Mathematics.
Franklin, C., Hartlaub, B., Peck, R., Scheaffer, R., Thiel, D., & Tranbarger Freiera, K. (2011). AP statistics: Building bridges between high school and college statistics education. *The American Statistician, 65*(3), 177–182.
Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*. Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/education/gaise/
Froelich, A. (2011). Developing a statistics curriculum for future secondary mathematics teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 27–32). New York: Springer.
Fry, E., & Garfield, J. (2015, January). *What do we know about best practices in teaching the introductory course?* Contributed paper presented at the Joint Mathematics Meetings, San Antonio, TX.
Fujii, Y., Fukazawa, H., Takeuchi, A., & Watanabe, M. (2014). A certification system for statistics knowledge and skills by Japanese Statistical Society. In K. Makar, B. de Sousa, &

R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_3B1_FUJII.pdf

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–25.

Gal, I. (2004). Statistical literacy—Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 47–78). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education, 2*(2). Retrieved from http://www.amstat.org/publications/jse/v2n2/gal.html.

Gapminder (2017, May 23). *Gapminder.* Retrieved from http://www.gapminder.org/.

Garfield, J. (1987). *Newsletter of the International Study Group for research on learning probability and statistics*, *1*(1). Retrieved from http://iase-web.org/documents/SERJIntStdGrp/studyv1n1.pdf.

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*(1), 25–34.

Garfield, J. (Ed.). (2005). *Innovations in teaching statistics*. Washington, DC: Mathematics Association of America.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372–396.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM: The International Journal on Mathematics Education, 44*(7), 883–898.

Garfield, J., delMas, R., Zieffler, A., & Fry, E. (2015, January). *e-ATLAS: Evaluation and assessment of teaching and learning about statistics.* Poster presented at the Joint Mathematics Meetings, San Antonio, TX.

Garfield, J., Hogg, B., Schau, C., & Whittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education, 10*(2), 1–15. Retrieved from http://www.amstat.org/publications/jse/v10n2/garfield.html

Gould, R. (2010). Statistics and the modern student. *International Statistical Review, 78*(2), 297–315.

Gould, R., Davis, G., Patel, R., & Esfandiari, M. (2010). Enhancing conceptual understanding with data driven labs. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics, Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_C208_GOULD.pdf

Green, D. R. (1982). *Probability concepts in school pupils aged 11–16 years* (Doctoral dissertation). Loughborough University of Technology, Loughborough, UK.

Green, J. L., & Blankenship, E. E. (2013). Primarily statistics: Developing an introductory statistics course for pre-service elementary teachers. *Journal of Statistics Education, 21*(3). Retrieved from https://ww2.amstat.org/publications/jse/v21n3/green.pdf.

Holmes, P. (2003). 50 years of statistics teaching in English schools: Some milestones. *The Statistician, 52*(Part 4), 439–474.

Hotelling, H. (1940). The teaching of statistics. *The Annals of Mathematical Statistics, 11*(4), 457–470.

Hunter, P. W. (1996). Drawing the boundaries: Mathematical statistics in 20th century America. *Historia Mathematica, 23*(1), 7–30.

Hunter, P. W. (1999). An unofficial community: American mathematical statisticians before 1935. *Annals of Science, 56*(1), 47–68.

Inman, H. F. (1990). The ASA section on statistical education: A historical note. *The American Statistician, 44*(2), 90–93.

International Association for Statistical Education. (2017a, May 23). *International statistical literacy project*. Retrieved from http://iase-web.org/islp/.

International Association for Statistical Education. (2017b, May 23). *Recent doctoral dissertations in statistics education.* Retrieved from http://iase-web.org/Publications.php?p=Dissertations.

Janssen, T., & Forbes, S. (2014). The use of official statistics in evidence-based policy making in New Zealand. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_5A1_FORBES.pdf

Jones, G. A. (Ed.). (2005). *Exploring probability in school: Challenges for teaching and learning*. New York: Springer.

Jones, P. S., & Coxford, A. F., Jr. (1970). Abortive reform—Depression and war: 1920–1945. In P. S. Jones (Ed.), *A history of mathematics education in the United States and Canada, 32nd Yearbook of the National Council of Teachers of Mathematics* (pp. 46–66). Washington, DC: National Council of Teachers of Mathematics.

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In F. K. Lester Jr. (Ed.), *The second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 909–955). Charlotte, NC: Information Age Publishing.

Kapadia, R., & Borovcnik, M. (Eds.). (1991). *Chance encounters: Probability in education*. Dordrecht, The Netherlands: Kluwer.

King, W. I. (1912). *The elements of statistical method*. New York: MacMillan.

Kudo, T., Watanabe, M., Morikawa, T., Iwasaki, M., Hayashi, Y., & Furutani, T. (2014). Challenges and issues in developing real-world curriculum for data scientists in Japan. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_1E2_KUDO.pdf

Langrall, C. W., Makar, K., Nilsson, P., & Shaughnessy, J. M. (2017). The teaching and learning of probability and statistics: An integrated perspective. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 490–525). National Council of Teachers of Mathematics: Reston, VA.

Lovett, M., Meyer, O., & Thille, C. (2008). The open learning initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. *Journal of Interactive Media in Education, 2008*(1). Retrieved from http://jime.open.ac.uk/articles/10.5334/2008-14/.

Loynes, R. M. (Ed.). (1987). *The training of statisticians round the world*. Voorburg, The Netherlands: International Statistical Institute.

Makar, K., & Ben-Zvi, D. (Eds.). (2011). The role of context in developing reasoning about informal statistical inference [Special issue]. Mathematical Thinking and Learning, 13(1–2).

Martignon, L. (2011). Future teachers' training in statistics: The situation in Germany. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 33–36). New York: Springer.

Michael, K., & O'Connell, A. (2014). Statistics education in Ethiopia: successes, challenges and opportunities. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_1D3_OCONNELL.pdf

Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, New Zealand: Learning Media.

Montgomery, D. C., & Woodall, W. H. (2008). An overview of Six Sigma. *International Statistical Review, 76*(3), 329–346.

Moore, D. S. (1988). Should mathematicians teach statistics (with discussion)? *College Mathematics Journal, 19*(1), 3–7.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123–165.

Moore, T. L. (Ed.). (2000). *Teaching statistics: Resources for undergraduate instructors*. Washington DC: Mathematics Association of America.

Moore, D., & Cobb, G. (2000). Statistics and mathematics: Tension and cooperation. *American Mathematical Monthly, 107*, 615–630.

Mosteller, F., Kruskal, W. H., Link, R. F., Pieters, R. S., & Rising, G. R. (Eds.). (1973). *Statistics by example (Volume 1: Exploring data, Volume 2: Weighing chances, Volume 3: Detecting patterns, Volume 4: Finding models)*. Reading, MA: Addison-Wesley.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2013). *TIMSS 2015 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Muñoz, E. C., Arañeda, A., Sorto, A., & León, J. L. (2014). Statistics and probability curriculum development for future elementary teachers in Chile: collaboration among countries. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/icots/9//proceedings/pdfs/ICOTS9_10C1_SORTO.pdf

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (1947). *Personnel and training problems created by the recent growth of applied statistics in the United States* (NRC Reprint & Circular Series, No. 128). Washington, DC: Author.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society—A, 97*, 558–606.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association, 33*(201), 101–116.

Neyman, J. (1976). The Emergence of mathematical statistics: A historical sketch with particular reference to the United States. In D. B. Owen (Ed.), *On the history of statistics and probability: Proceedings of a symposium on the American mathematical heritage to celebrate the bicentennial of the United States of America* (pp. 149–193). New York: Marcel Dekker.

OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, and problem solving knowledge and skills*. Paris: OECD Publishing. https://doi.org/10.1787/9789264190511-en

Opolot-Okurut, C., & Eluk, P. (2011). Statistics school curricula for Uganda. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 15–20). New York: Springer.

Ottaviani, M. G. (2000). Research into statistical education. The point of view of a statistician. *Statistical Education Research Newsletter, 1*(1).

Parsian, A., & Rejali, A. (2011). An experience on training mathematics teachers for teaching statistics in Iran. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 37–40). New York: Springer.

Pearson Committee. (1947). Report on the teaching of statistics in universities and university colleges. *Journal of the Royal Statistical Society, 110*(1), 51–57.

Pearson, E. S., Wishart, J., Barnard, G. A., Bartlett, M. S., Champernowne, D. G., & Finney, D. J. (1955). A discussion on the teaching of mathematical statistics at university level. *Journal of the Royal Statistical Society, Series A (General), 118*(2), 193–223.

Peng, R. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance, 12*(3), 30–32.

Perry, J. (1900). The teaching of mathematics. *Nature, 2*, 317–320.

Pfannkuch, M., Forbes, S., Harraway, J., Budgett, S., & Wild, C. (2013). *"Bootstrapping" students' understanding of statistical inference*. Summary research report for the Teaching and Learning Research Initiative. Retrieved from https://www.tlri.org.nz/sites/default/files/projects/9295_summary%20report_0.pdf.

Porkess, R. (2011). *The future of statistics within our schools and colleges*. London: Royal Statistical Society.

Radford, L. (Ed.). (2015). Statistical reasoning: Learning to reason from samples [Special issue]. Educational Studies in Mathematics, 88(3).

Randles, R. H. (2007). A conversation with Robert V. Hogg. *Statistical Science, 22*(1), 137–152.

Reston, E., & Bersales, L. G. (2011). Reform efforts in training mathematics teachers to teach statistics: Challenges and prospects. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 41–46). New York: Springer.

Reston, E., & Jala, L. L. (2014). Sustaining teachers' capacity for teaching statistical inference through reflective practice. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_1D4_RESTON.pdf

Rice, S. A. (1949). Furtherance of statistical education: Report to the XXVIth Session of the International Statistical Institute. *International Statistical Review, 17*(1/2), 1–34.

Richardson, A. (2014). Building capability in statistics majors: Drawing strength from a diverse region. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_1E3_RICHARDSON.pdf

Roberts, R., Scheaffer, R., & Watkins, A. (1999). Advanced placement statistics—Past, present, and future. *The American Statistician, 53*(4), 307–320.

Rossman, A., & Chance, B. (2013). *Workshop statistics: Discovery with data and minitab*. New York: Springer.

Rossman, A., & Garfield, J. (2011). Interview with Joan Garfield. *Journal of Statistics Education, 19*(3). Retrieved from http://www.amstat.org/publications/jse/v19n3/rossmanint.pdf.

Royal Statistical Society. (1947). Report on the teaching of statistics in universities and university colleges. *Journal of the Royal Statistical Society A, 110*, 51–57.

Saxe, K., & Braddy, L. (2015). *A common vision for undergraduate mathematical sciences programs in 2025*. Washington, DC: Mathematical Association of America.

Schau, C., Stevens, J., Dauphinee, T., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes Toward Statistics. *Educational & Psychological Measurement, 55*(5), 868–876.

Scheaffer, R. L. (1990). The ASA–NCTM quantitative literacy project: An overview. In D. Vere-Jones, S. Carlyle, & B. P. Dawkins (Eds.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 1–5). Voorburg, The Netherlands: International Statistical Institute.

Scheaffer, R. L. (2001). Statistics education: Perusing the past, embracing the present, and charting the future. *Newsletter for the Section on Statistical Education, 7*(1). Retrieved from http://www.amstat.org/sections/educ/newsletter/v7n1/Perusing.html.

Scheaffer, R. L. (2003). Statistics and quantitative literacy. In B. L. Madison & L. A. Steen (Eds.), *Quantitative literacy. Why numeracy matters for schools and colleges* (pp. 145–152). Princeton, NJ: The National Council on Education and the Disciplines.

Scheaffer, R. L., & Jacobbe, T. (2014). Statistics education in the K-12 schools of the United States: A brief history. *Journal of Statistics Education, 22*(2). Retrieved from http://www.amstat.org/publications/jse/v22n2/scheaffer.pdf.

Schuyten, G., & Ottaviani, M. (2006). Fifteen years of IASE: Mission and instruments. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the 7th International Conference on Teaching Statistics, Salvador, Bahia, Brazil [CD-ROM]*. Voorburg, The Netherlands: International Association for Statistical Education and International Statistical Institute.

Shaughnessy, J. M. (1991). Research in probability and statistics reflections and directions. In D. Grouws (Ed.), *Handbook for research in mathematics teaching and learning* (pp. 465–494). New York: Macmillan Publishing.

Shaughnessy, J. M. (2006). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second handbook for research in mathematics teaching and learning* (pp. 957–1009). Greenwich, CT: Information Age Publishing and NCTM.

Shi-Jian, Y. (1990). Probability and statistics courses in the universities in China. In D. Vere-Jones, S. Carlyle, & B. P. Dawkins (Eds.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 85–90). Voorburg, The Netherlands: International Statistical Institute.

Shubring, G. (2013). The road not taken—The failure of experimental pedagogy at the Royaumont Seminar 1959. *Journal für Mathematik-Didaktik, 35*(1), 159–171.

Shulte, A. P., & Smart, J. R. (Eds.). (1981). *Teaching statistics and probability: 1981 yearbook*. Reston, VA: National Council of Teachers of Mathematics.

Smith, A. (2004). *Making mathematics count: The report of professor Adrian Smith's inquiry into post-14 mathematics education*. London: Her Majesty's Stationery Office. Retrieved from http://dera.ioe.ac.uk/4873/1/MathsInquiryFinalReport.pdf

Snedecor, G. W. (1948). A proposed basic course in statistics. *Journal of the American Statistical Association, 43*(241), 53–60.

Sorto, M. (2011). Statistical training of Central American teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 47–51). New York: Springer.

SRTL. (2017). Research forums. *International collaboration for research on statistical reasoning, thinking and literacy* [website]. Retrieved from srtl.info.

Stigler, S. M. (1996). The history of statistics in 1933. *Statistical Science, 11*(3), 244–252.

Swanstrom, R. (2017, May 23). *College & university data science degrees*. Retrieved from http://datascience.community/colleges.

Tanur, J., Mosteller, F., Kruskal, W. H., Link, R. F., Pieters, R. S., & Rising, G. R. (1972). *Statistics: A guide to the unknown*. San Francisco: Holden-Day.

Tintle, N., Chance, B., Cobb, G., Rossman, A., Roy, S., & VanderStoep, J. (2014). *Introduction to statistical investigations*. Hoboken, NJ: Wiley.

Tintle, N., Chance, B., Cobb, G., Roy, S., Swanson, T., & VanderStoep, J. (2015). Combating anti-statistical thinking using simulation-based methods throughout the undergraduate curriculum. *The American Statistician, 69*(4), 362–370.

Tolley, H. R., & Ezekiel, M. J. B. (1923). A method of handling multiple correlation problems. *Journal of the American Statistical Association, 18*(144), 993–1003.

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics, 33*(1), 1–67.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Tulya-Muhika, S. (1990). Teaching statistics for future government statistical services in Africa. In D. Vere-Jones, S. Carlyle, & B. P. Dawkins (Eds.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 552–561). Voorburg, The Netherlands: International Statistical Institute.

Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician, 57*(2), 74–79.

Utts, J. (2010). Unintentional lies in the media: Don't blame journalists for what we don't teach. In C. Reading (Ed.), *Data and context in statistics education: Proceedings of the Eighth International Conference on Teaching Statistics*. International Statistical Institute: Voorsburg, The Netherlands. Retrieved from https://iase-web.org/documents/papers/icots8/ICOTS8_1G2_UTTS.pdf

Utts, J. (2015). The many facets of statistics education: 175 years of common themes. *The American Statistician, 69*(2), 100–107.

Vere-Jones, D. (1995). The coming of age of statistical education. *International Statistical Review, 63*(1), 3–23.

Walker, F. A. (1890). Study of statistics in colleges and technical schools. *Technology Quarterly, 3*(February), 1–8.

Walker, H. M. (1929). Statistics as a subject of instruction in American universities. In *Studies in the history of statistical method with special reference to certain educational problems* (pp. 148–174). Baltimore, MD: The Williams and Wilkins Company.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on *p*-values: Context, process, and purpose. *The American Statistician, 70*(2), 129–133.
Watson, J. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum Associates.
Watson, J. (2014). Curriculum expectations for teaching science and statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_1A1_WATSON.pdf
Welch, B. L. (1935). Some problems in the analysis of regression among *k* samples of two variables. *Biometrika, 27*(1–2), 145–160.
Wessels, H. (2011). Statistics in the South African school curriculum. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 21–26). New York: Springer.
West, W. (2014). Teaching resampling in an introductory statistics course. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_9B3_WEST.pdf
Wikipedia Contributors. (2016, February 20). Statistics education. In *Wikipedia, the free encyclopedia*. Retrieved July 3, 2017, from https://en.wikipedia.org/wiki/Statistics_education.
Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–265.
Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 174*(2), 247–295.
Wild, C. J., Pfannkuch, M., Regan, M., & Parsonage, R. (2017). Accessible conceptions of statistical inference: Pulling ourselves up by the bootstraps. *International Statistical Review, 85*(1), 84–107.
Willcox, W. F. (1910). The outlook for American statistics. *Publications of the American Statistical Association, 12*(89), 43–51.
Wilson, E. B. (1930). Mathematics and statistics. *Journal of the American Statistical Association, 25*(169), 1–8.
Wilson, S. (2013). The flipped class: A method to address the challenges of an undergraduate statistics course. *Teaching of Psychology, 40*(3), 193–199.
Winquist, J., & Carlson, K. (2014). Flipped statistics class results: Better performance than lecture over one year later. *Journal of Statistics Education, 22*(3). Retrieved from http://www.amstat.org/publications/jse/v22n3/winquist.pdf.
Wishart, J. (1939). Some aspects of the teaching of statistics (with discussion). *Journal of the Royal Statistical Society A, 102*, 532–564.
Woodard, R. & McGowan, H. (2012). Redesigning a large introductory course to incorporate the GAISE guidelines. *Journal of Statistics Education, 20*(3). Retrieved from http://ww2.amstat.org/publications/jse/v20n3/woodard.pdf.
Yates, F. (1946). A review of recent statistical developments in sampling and sampling surveys. *Journal of the Royal Statistical Society, 109*(1), 12–43.
Zeleke, A., Lee, C., & Daniels, J. (2006). Developing projects based on student's data in introductory statistics. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS-7, July, 2006), Salvador, Brazil*. Voorburg, The Netherlands: ISI.
Zieffler, A., & Catalysts for Change. (2013). *Statistical thinking: A simulation approach to uncertainty* (3rd ed.). Minneapolis, MN: Catalyst Press.
Zieffler, A., & Fry, E. (Eds.). (2015). *Reasoning about uncertainty: Learning and teaching informal inferential reasoning*. Minneapolis, MN: Catalyst Press.

Chapter 3
Statistics Education Research

Peter Petocz, Anna Reid, and Iddo Gal

Abstract This chapter sketches in broad strokes and critically examines several aspects of the world of research that pertain to the teaching, learning, understanding, and using of statistics and probability in diverse contexts, both formal and informal. It reflects on the methods and conceptual schemes that underlie the research activity in this field (the *how*), the topics being researched (the *what*), and the people carrying out the research (the *who*). The chapter examines purposes and motivations for different types of studies in statistics education, distinguishing between *large-R* research that often aims for academic reporting and generalizability versus *small-r* types of research whose motivation is more on local problems set in a particular context. We illustrate some trends in the field by presenting empirical results from an exploratory qualitative analysis of the text of a body of papers and publications in the field. The chapter points out that the range of what qualifies as research in (or of relevance to) statistics education is much broader than what gets published in leading journals and conferences in our field. It highlights the multiplicity of philosophical foundations and methodologies in use. Some directions for future development and research are outlined, including aspects of statistical literacy, cultural dimensions of statistics education research, the role of practitioner inquiry, and the importance of broad interdisciplinary research in statistics education.

Keywords Statistics education • Research methods • *Small-r* and *large-R* research • History of research • Practitioner research • Statistical reasoning • Statistical literacy • Research utilization

---

P. Petocz (✉)
Department of Statistics, Macquarie University, North Ryde, NSW, Australia
e-mail: Peter.Petocz@mq.edu.au

A. Reid
Sydney Conservatorium of Music, University of Sydney, Sydney, NSW, Australia
e-mail: Anna.Reid@sydney.edu.au

I. Gal
Department of Human Services, The University of Haifa, Haifa, Israel
e-mail: iddo@research.haifa.ac.il

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_3

3.1 Introduction

In this chapter, we take a broad overview of the field of research of relevance to statistics education in order to highlight some overall trends and approaches. We aim to reflect, albeit in a somewhat personal way, on three broad questions related to the *how*, *what*, and *who* of statistics education research, that is, the methods and conceptual schemes that underlie some of the research activity in this field, the topics or issues being researched, and the people carrying out the research.

Despite the diversity in the identity of the researchers, or the targets of their curiosity and motivation, research work is of relevance to us (i.e. to the area of statistics education) if it has the potential to contribute to the field of statistics education and enhance in some way the ability of learners or adults to think or reason about statistical or probabilistic situations or act in a statistically literate way. This is enabled by the fact that *what* is being investigated relates in some way to the basic building blocks around which teaching, learning, and using statistics and probability revolve. After all, there are common features stemming from the big ideas and core concepts that make statistics a unique field (see Chap. 1, this volume) and (should) permeate all levels and contexts of instruction or usage (see Chap. 2, this volume). This is the case, for instance, for ideas related to variability and distribution; uncertainty, probability, error, and risk; sampling and generalization; modelling and data reduction; inference, prediction, and causality; existence of methods for exploring and representing notions of distribution, centre, dispersion, or association in diverse ways such as in via texts, tables, numbers, or graphs and dynamic visual displays; and many more (see, e.g. Moore, 1990; Moore & Cobb, 2000, for additional examples).

The idea of *what* is being investigated is powerful, as a clear definition of *what* leads intuitively and practically to *how* it could be explored. A particular feature of statistics education research is that the disciplinary content or the pedagogical context may also inform the method deemed suitable for investigation. In some contexts, for some people, it may seem very natural to investigate aspects of learning statistics using a statistical (quantitative) approach. The extension of this idea is that the method of investigation is predetermined and so needs no discussion, or even the view that any alternative method of investigation is in some way inferior to the statistical method. Pedagogical research in (say) chemistry may tend to favour a quantitative approach, and in (say) history may tend to favour a qualitative approach. However, in no discipline other than statistics is there such an immediate link between the subject matter under investigation and the methodology that may seem appropriate for the investigation—even if that subject matter is statistics *education*. That said, the actual range of methodologies used in research of relevance to statistics education is broad and goes well beyond those normally used by professional statisticians who engage in the practice of statistics.

This chapter is organized in several sections. We start by reiterating, and critically examining, a distinction between the hallmarks that identify educational investigations as reportable research (we refer to these as '*large-R* research') and other forms (which we refer to as '*small-r* research') that are not primarily aimed at publication as research, though they may be described in journals or conferences focusing on statistics education. In putting forward this distinction, it is not our intention to valorize the *large-R* research or to disparage the *small-r* research. On the contrary, we believe that both types have an important role to play in statistics education research generally, and their combination results in a more detailed overall picture than would be obtained following the more conventional approach of excluding the *small-r* results on the basis that they do not meet some criterion of research quality.

To inform some of our ideas, we report selected results from an exploratory analysis of a corpus of contemporary publications in the field of statistics education research, using the qualitative research package NVivo. Following a brief overview of this investigation, we next provide a general discussion of research philosophies (too often omitted in written descriptions of research results; see also Chap. 11 on theories in statistics education research, this volume), including an example of the multiplicity of these perspectives in the context of research on statistical literacy. With the above as a backdrop, in the remainder of this chapter, we continue to look closely at the three linked aspects, the *how*, the *what*, and the *who* of research in statistics education. Our approach is based on our conceptual investigation of the topics of statistics education research on the one hand and our empirical investigation of the artefacts of statistics education research on the other. We also discuss *why* some of the objects of research have been selected and the way in which the selected approaches fit particular learning, social, and cultural situations. We then close with discussion and conclusions.

The field of statistics education research provides us with an opportunity to explore the value of research approaches that are typically used by statistics educators and understand how this field is unique and at the same time generic. Overall, the chapter aims to inform beginning researchers and scholars, yet also to be of interest to more seasoned researchers, by offering a reflective yet critical assessment regarding the evolution of the field of statistics education research.

3.2 Background

3.2.1 Distinguishing Types of Research

Evans (2010) gives a useful guide to what can be seen to as academic research. He points out that the practice must be integrated within a strategy that is purposive (based on the identification of an issue or problem worthy and capable of investigation), inquisitive (seeking to acquire new knowledge), informed (conducted from an awareness of previous, related research), methodical (planned and carried out in a disciplined manner), and communicable (generating and reporting results which are testable and accessible by others). Evans' guide highlights characteristics of research that can well be applied to the field of statistics education research. Purposive research allows us to situate the research in the societal context of the problem—for instance, some research is situated in the context of ensuring that, as part of the final outcome of schooling, future citizens and workers are able to access and understand statistical information presented numerically or graphically, whereas other research is situated in the context of the goal of developing formal knowledge and technical skills of future professional statisticians. Purposive research is based on knowledge of the field and the selection of appropriate methods to perform the investigation. An outcome of purposive research is that new knowledge should be acquired and promulgated. The nature of the problem and its identification also demonstrate the epistemological position that the researchers have taken, and the epistemological position generates the research approach.

That said, although both *large-R* and *small-r* research may be purposeful, one of the purposes of most *large-R* research is generalization, whereas *small-r* research more commonly (though not always) focuses on localized issues. Thus, *small-r* researchers may not necessarily have an explicit intention to generalize results and publish them to a wider community. Examples are the case of a practitioner who studies their own practice, an official statistics agency exploring the degree to which users understand a certain statistical product (e.g. consumer price index) and are satisfied with the explanations given on an agency website, or an internal formative evaluation process that is conducted as part of funded research that involves learning or using statistics, such as a project in the area of science education.

Another key common thread between both the *small-r* and *large-R* research problems is curiosity. Such curiosity motivates the people *who* carry out the research, and at the same time it defines an object of research. For instance,

- Educators may become curious about the way in which children develop statistical understanding or about the impact of factors such as family conditions, curriculum materials, type of instruction, or use of technology on learners' comprehension.
- Lecturers may be interested in how the attitudes or backgrounds of tertiary (or college) students affect their performance, how to improve the teaching-learning process, or students' ability to apply their knowledge in professional situations.
- Managers may notice what statistical messages or risk information may be misunderstood by workers or service recipients and wonder how to improve comprehension.
- Researchers in a certain discipline may be curious about the factors affecting the statistical thinking and behaviour of their respondents.

The curiosity on the part of teachers, professionals, or researchers is what initiates and enables research of relevance to statistics education. In some cases, educators may decide to explore something in their classroom that could make an immediate impact for a specific group of students. In others, broader conceptual and practical problems generate a prolonged investigation of some rather complex learning, cultural, and statistical issues. Selected examples can be found in many of the chapters in this handbook, which explore in depth some objects of research that have been established as critical concepts in statistics education, often highlighting the scholars or research groups who are carrying out the research. Other chapters explore some of the common research methods and approaches that statistics educators use in their investigations of particular areas or topics.

3.2.2 Distinguishing Types of Researchers

In this chapter, we call the people involved in statistics education research for brevity 'statistics education researchers', though this term is simplistic. In fact, there are very diverse types of people and groups involved in research in statistics education. Here are key examples (though others do exist); some may have commonalities or overlap, but overall they illustrate many differences:

(a) Traditional forms of research that focus on statistics education processes or outcomes in formal education contexts such as schools or tertiary institutions
(b) Localized practitioner-based research in which a single teacher or a group within a single school decide to examine their own practice in a systematic way, not necessarily with the hope of generalizing
(c) Broader practitioner-based research, possibly by teachers in academia or at a graduate level, who collaborate across multiple sites on studying the reactions to or benefits from a certain innovation such as a new teaching method, a new assessment system, or a digital application
(d) Practical research conducted by producers of statistics (e.g. official statistics agencies) or educational actors who operate outside formal education contexts (e.g. trainers needing to improve how workers understand data on quality of manufacturing or service processes), aiming to inform localized decision-making
(e) Formative research about ongoing educational interventions or applied projects that involve some statistics, such as programme evaluations or surveys of students and/or teachers in the area, aiming to inform further planning and execution of the programme
(f) Research conducted by scholars from other disciplines who investigate a specific issue of relevance to the learning or teaching of statistics or probability yet do not necessarily have statistics education as their main focus but instead aim to inform issues of interest in another discipline such as mathematics education, science education, psychology of judgement and decision-making or risk communication, health education, or educational technology

3.2.3 Surveying the Landscape: An Empirical Analysis

How does one go about investigating the *how*, *what*, and *who* aspects that underlie the research activity in statistics education? Other chapters in this volume each focus on a single selected subarea and explore insights from research results in this regard. In this chapter, however, our aim has been to provide an overview of the whole field regarding the *how*, *what*, and *who* questions, going above and beyond selected subareas, as important as they are.

Using the qualitative research package NVivo 10, we have investigated a corpus of contemporary publications in the field of statistics education research. Since we wrote the first version of this chapter in 2015, we chose to base our analysis on a half-decade of articles published from 2010 to 2014 inclusive. We collected all the papers published in the *Statistics Education Research Journal* (*SERJ*), the *Journal of Statistics Education* (*JSE*), and the *Technological Innovations in Statistics Education* (*TISE*). We also included the 'invited' (though not the 'contributed') papers presented at two recent *International Conferences on Teaching Statistics* (ICOTS 8 in 2010 and ICOTS 9 in 2014) and published on their conference websites and papers from several Satellite Conferences and Roundtables organized by the *International Association for Statistical Education* (IASE) and available via its website: http://iase-web.org/Conference_Proceedings.php.

This approach led to a collection of 653 papers: a body of papers not without limitations but also with some advantages. Firstly, it includes those that deal with various aspects of statistics education as well as statistics education research, as we did not go through a process of selecting only those that were focused on research. That way, we intentionally kept the *small-r* research papers as well as the *large-R* research papers, since we believe that both types are important. The inclusion of proceedings from two ICOTS conferences in a 5-year period biases the selection towards conference papers, though at the same time it highlights the international diversity and practitioner-led research that are amongst the features of these key conferences. Further, the sources listed above are not the only places where research in statistics education is published; for instance, we have not examined the (much larger) body of research in mathematics education to identify studies focusing on statistics education. A search through several mathematics education research journals for the period 2010–2014 revealed only a small number of papers on statistics education research. In seven leading journals we found a total of 20 such papers, including nine in a special issue of a single journal (see Makar & Ben-Zvi, 2011). Despite their important disciplinary contribution, the actual number of such papers is too small to have any noticeable effect on the overall results of this analysis. Although the coverage of our collection of articles is not complete, it does represent a coherent and inclusive corpus of literature for investigation.

NVivo allowed us to search through the complete text of this collection of articles using a variety of search terms derived initially from the combined contents of all the papers and focused specifically on the aspects of research approach and method. Despite its limitations, our analysis was purposeful. We tried to illustrate what is found within a sensible but constrained search space, in order to inform our own writing of this chapter. The collection of 653 articles gives a comprehensive view of the world of discourse in statistics education and statistics education research over a recent half-decade—a view that has not been previously reported in the literature on statistics education.

3.3 Defining the Philosophies

Ontology, or ‘ways of being’, and epistemology, or ‘ways of knowing’, are at the core of all research. A paradigm is constituted of ontology, epistemology, method, forms of analysis, discipline of research, and interpretation. An ontological perspective focuses the researcher’s attention on the essence of something or the existence of something—for instance, the essential meaning of variability or the nature of a statistical approach to a problem or to life in general. An epistemological perspective seeks to understand how people come to know something—for instance, how they learn about different types of averages or the discipline of statistics as a whole. Researchers need to consider the nature of the thing that they are investigating, and understanding the interrelationships between the ontology and epistemology provides a powerful tool that guides our research practice. For instance, Petocz and Reid (2010) highlight the importance of the ontological aspect of *becoming* (and *being*) a statistician, in addition to the epistemological aspect of learning *about* statistics, in the process of statistics education. Many researchers, however, are unaware of these distinctions and take a rather more naturalistic approach to their investigations, and this can account in some cases for the distinction between *small-r* and *large-R* activity. Statistics education researchers need to be aware of their assumptions about the nature of statistical knowledge so that they can appropriately carry out their investigations in ways that are reliable and valid. To start to explore this idea, we will think about some extremes in the nature of knowledge. For statisticians, positivism is an idea that is core to the discipline.

3.3.1 Positivism

Scotland (2012) suggests that positivism has its basis in realism. That is, things exist and have an intrinsic value, and the role of the researcher is to find that intrinsic value and describe it. Hence, data that are factual in nature reside in this realm. Ways of exploring this knowledge create an epistemology that is founded on the collection of objective facts and the provision of an analysis that is about the object of research and therefore free of cultural subjectivity. Some aspects of this view of research were challenged by the post-positivists (see Popper, 1959), who pointed out that every question is determined by a person, and every finding can be critiqued or questioned. What could be considered objective in an ontological sense can also be considered subjective. This perception brings into view the notion that a researcher and research question are part of a social construction of knowledge and curiosity. A mountain can still exist without a person thinking about it, but the moment that they do think about the mountain, then they bring to the thought their own experience and social history.

The main strength of a positivist (and post-positivist) approach to research is that it seeks to find testable causes for a situation that can be explained through relationships between things. For statistics education researchers, and researchers in education generally, this can result in definite answers to specific questions, in a form that is aligned with standard scientific approaches and methods, and can satisfy the requirements of funding agencies. Whole research programmes can be set up using such a positivist framework. For instance, members of the *American Statistical Association* set forth guidelines for research in statistics education for introductory college-level statistics. Pearl et al. (2012) gives a comprehensive listing of research classified into six main areas, with research priorities and questions for each area, and a short statement of the benefits that would result from knowing the answers to these questions.

Methodologies that are associated with this form of research include survey instruments, experiments with controls, standardized tests, etc. Analysis of the data that come from these methods includes descriptive or inferential statistics that can be generalized to broad populations. An important aspect of this form of research is that it can be replicated to produce results that should be similar. For instance, a researcher may want to explore how students react to a different teaching method designed to help students understand probability. A pretest could be administered to two classes to gauge their current knowledge of probability. One class could have the different teaching method applied and the other class not. In the end, a post-test will determine which group actually understood probability in a better way than the other.

A detailed description of such an approach is presented in the SMER Report 'Using Statistics Effectively in Mathematics Education Research' (ASA, 2007). However, as that Report points out, for statistics education researchers, a strict scientific approach can be problematic. For example, in a learning environment, controlled conditions are rarely possible. Students are allocated to classrooms or lecture groups in a non-random way; they have learning experiences that include interaction with students from other classes and with other materials; 'treatment' may only be possible class by class, and the notion of single or double blind is usually completely unfeasible. Unlike other experimental designs in some other areas, it is nearly impossible to replicate educational studies in other learning environments. Statistics education researchers who base their research on a positivist approach need to be aware of the inherent limitations.

The approach taken here is familiar from the broader scientific context, for instance, a clinical trial to investigate the effect of a drug versus a placebo. Aspects of such a clinical trial that are considered important scientifically include random assignment to treatment or placebo group, (single or double) blinding to avoid bias on the part of the subjects and the experimenters, and the possibility of replicating the trial with another group of subjects. This produces the so-called gold standard of scientific research that is held up as ideal in some contexts. There are numerous examples for expectations or standards in this regard (e.g. Shelley, 2005 discusses the US Government's 2001 legislation to foster scientifically valid research in educational research).

Hill and Shih (2009) have examined the quality of research in mathematics education, based on an analysis of 10 years of articles in the *Journal for Research in Mathematics Education* (*JRME*). Using criteria suggested by several well-known professional associations, such as the *American Educational Research Association*, *American Psychological Association*, and *National Council for Measurement in Education*, or the *American Statistical Association*'s report on using statistics in mathematics education research (ASA, 2007), these authors have found that the majority of *JRME* articles reporting on quantitative research were lacking in one or more respects.

3.3.2 Social Constructivism

A completely different way of thinking about knowledge is that it is constructed socially. The ontological position is that knowledge is relative and open to interpretation. Relativism is an idea that acknowledges the different ways that reality is perceived from person to person (Guba & Lincoln, 1994, p. 110). Unlike positivism, where things have an intrinsic quality, a socially constructed view indicates that it has been given meaning through a person's interaction with it and other things. On the whole, socially constructed knowledge can be seen as subjective rather than objective as in the positivist paradigms. This way of viewing things brings with it quite a different epistemology. Now, ways of knowing are based on both individual and social experiences. All people experience things, and experience things uniquely. Individual life experiences are unique and cannot be replicated across cultures and time. However, it is a characteristic of people to be curious and intentionally seek to understand and experience things in a different way. Small children notice variation in the way that adults make sounds and how it differs from their own attempts. They notice that their mouths and tongues move in different ways when they try to replicate those sounds, eventually they notice that the sounds have order and create meaning, and finally they learn to speak. In every human enterprise, noticing variation is the way that we learn.

Statistics education researchers are experts at noticing variation. It is at the core of their discipline and usually encountered using more positivist activities. When it comes to understanding learning, teachers notice immediately if there are more girls than boys in their class, or if the students are young or old, or dark- or fair-haired, or whether they are local students or come from another country or background. These are easy observations to make, and they assist the teachers to understand the social situation in which they find themselves. However, it is harder to notice variation in the way that people think. It is here that statistics education researchers are often curious. To explore aspects of their students' learning, educators need to use methods that enable students (or participants) to explain their experience. Interview studies, case studies, talk-aloud protocols, narrative inquiry, phenomenology, etc. all fit into this paradigm. Where these (and other) methods differ is in their epistemology—the way in which knowledge is formed. Further discussion on such issues and illustrations regarding the use of qualitative methods to inform statistics education can be found in Petocz and Reid (2010), Gal and Ograjenšek (2010), and many other sources.

Taking a narrative inquiry approach, a statistics education researcher may look at curriculum documents, text books, exam papers, and student diaries to come to an understanding of how students experience probability. Taking this approach, the researchers can investigate why curriculum authors think it is important for students to learn about probability, how it is presented in textual material, how it is made appropriate to different age groups and cultures, how students experience it inside and outside the classroom, and how students may start to use probability as a way of thinking in daily life or as part of a more scientific enterprise. The intended outcome of the research is knowledge regarding student experience, with the intention of finding more effective ways of working with such students.

A phenomenological approach would be distinctly different from narrative inquiry. Such an approach aims at providing a rich description of an individual's experience of probability (for instance). The researcher may use interviews, videos, discussion groups, observations, diaries, blogs, etc. The purpose of the data collection is to provide a 'thick description' of the individual experience. The outcome is to use the description to understand other similar individuals and situations through noting variation and similarity. A key characteristic of the approach is that it focuses on the individual as a creator of knowledge and legitimizes the authority of that knowledge.

3.3.3 Critical Theory

Critical theory provides an alternative ontology and epistemology that can be of use to the statistics education researcher. This theory extends beyond positivist as well as interpretive approaches, as it takes account of the history, politics, and society in which people create knowledge. Knowledge is seen as inherently political as it reflects the values that people place on it. From this perspective, epistemology includes gender (feminism, queer theory), culture (Marxism, Confucianism), liberty (self-determination, poverty), critical economics or peace studies, etc. Scotland (2012, p. 11) says:

> As it is culturally derived, historically situated and influenced by political ideology, knowledge is not value free. The critical paradigm asks the axiological question: what is intrinsically worthwhile? Thus, the critical paradigm is normative; it considers how things ought to be; it judges reality. The utopian aspirations of the critical paradigm may never be realized but a more democratic society may materialize.

Critical theory can be applied to both positivist and constructivist paradigms. An example within the (post-)positive paradigm could be the graphical investigations on the Gapminder website (see www.gapminder.org) of various aspects of contemporary society, for instance, the relationship between education of girls/women and population growth. The graphs demonstrate that in countries where a greater proportion of girls are educated, there is a lower rate of births per woman, and this relationship can be visualized over time. An example within the constructivist paradigm could be Vita and Kataoka's (2014) investigation of modifying teaching sequences and materials to help blind students in their learning of probability. Their study illustrates their acknowledgement of their students' diversity and their ethical concern to provide learning opportunities for them that are appropriate to this diversity. Several other articles, and maybe even the complete special issue, in this volume of the *Statistics Education Research Journal* (*SERJ*, 13(2)) highlight aspects of critical theory.

3.3.4 Example: Applying Different Approaches to Statistical Literacy

The idea that knowledge based on research results (including from quantitative research) is socially constructed, and that statistical statements made by researchers or statistics producers are legitimate targets for a critical process, underlies the motivation to develop the statistical literacy of future and current citizens (Gal, 2002a). This is now evident in many curriculum frameworks around the world. However, research on statistical literacy by itself can take different forms, both positivistic and interpretive.

Some researchers in the area of statistical literacy focus on obtaining quantitative or quantifiable evidence regarding ability to critically interpret statistical messages, using quantitative methods such as tests with multiple-choice items or by employing rubrics for coding responses to open-ended tasks (e.g. ability to recognize flaws in a media article with statistics). Numerous examples for such approaches appear in *SERJ* and other sources we analysed, for instance, Hobden's (2014) work on the level of statistical literacy proficiency of pre-service teachers in understanding statistics about HIV/AIDS in South Africa. Such quantitative studies, and theoretical models on which they are based (e.g. Watson & Callingham, 2003), are valuable in helping to document ability levels of different target groups of interest or the impact of using certain interventions.

However, researching the *inner change* process that learners' undergo as they develop the 'critical lens' that is part of statistical literacy requires different and less direct approaches. Action research, critical discourse analysis, phenomenography, and some other approaches that use open-ended questioning techniques or 'think aloud' cognitive protocols may be useful in this regard. The need for such methods is in part due to the need to simultaneously document changes in both cognitive and attitudinal systems (Gal, 2002b). In addition, the critical perspective being adopted by the learners, by its very nature, may be perceived differently by the researcher and by those being researched (i.e. the learners undergoing the change). Hence, to understand how statistical literacy develops, there is a need to include participants as co-researchers, recognizing that researchers *and* participants provide equally legitimate contributions to the research object. This in turn implies a need to work with relatively mature populations of participants, who can take part in a co-creation process where they are both learners and collaborators in a reflective process with the researchers. An example is recent work by Brantlinger (2014), who engaged adult students taking a night class in the critical interpretation of a chart portraying the white student to student-of-colour ratio at various schools and how these correlated with recess time received by students. The participants not only reflected on the value of this activity as an educational task but also on the extent to which it enables the researcher to understand their own perspectives.

In moving from positivist or socially constructed to critical theoretical approaches, the role of a considered stance to the ethical aspects of research becomes more prominent; this is as true in statistics education research as in any other area. There are ethical aspects to any research, even if the epistemology is scientific and positivist. Yet in that case, the search for the truth is often accepted as the most important feature of the research, and any problematic aspects related to the 'subjects' or even the experimenters themselves are often seen to be secondary. With the socially constructed epistemologies, the role of the people involved in the investigations becomes more central. The ethical aspects of how they are treated and how their evidence is used become an integral part of the research. Appreciating the ethical aspects of research enables the researcher to work sympathetically and carefully with any subject/participant group. In the statistics education context, this is important as, for instance, any group of students that is involved in research is also likely to be the recipient of the outcomes of the research, as some aspect of their learning situation is likely to change. The critical epistemologies problematize the ethical dimension further, and include consideration of the possible negative consequences of participation in research projects, and the ultimate ownership of the resulting research, including the right to veto any further use of material obtained and even knowledge uncovered.

3.4 Reflections and Evidence on the *How*, the *What*, and the *Who*

3.4.1 Methods Commonly Used by Statistics Education Researchers (the How*)*

In this section we will investigate the range of methods, approaches, and methodologies used by contemporary statistics education researchers. The body of evidence is the 653 articles from *SERJ*, *JSE*, and *TISE* and the invited papers published on the websites of the ICOTS 8 and 9 and IASE satellite and round table conferences from the period 2010–2014. These articles were selected to cover a broad range of writings in the field of statistics education research, published in the highest profile journals and conferences in the field, and included *small-r* as well as *large-R* research writing. The selection makes no claims to being complete (though we believe that it is quite comprehensive), nor to being a random sample, nor was it weighted in any way to represent issues, researchers, or quality; the results should be read with this in mind.

NVivo enabled us to investigate references to a variety of methodological approaches using a number of search terms. These terms were selected from an initial examination of the full text of all the articles based on our aim to examine their research methods and approaches. The following search terms were selected and used:

General: research approach, research method(ology), ontology, epistemology.

Quantitative approaches: questionnaire, factor analysis, structural equation model.

Qualitative approaches: content analysis, constructivism, focus group, action research, lesson study, content knowledge, interview (study), reflective practice, phenomenology.

Other terms such as regression, hypothesis test, and analysis of variance could also be taken to indicate quantitative research approaches, but investigation of the articles that used these terms revealed that they were overwhelmingly references to statistical techniques that were being taught and/or learned; this was not the case for the other quantitative terms.

The profile of terms and the number of papers in which they were used indicated that the majority of the articles did not make explicit reference (using our terms) to philosophical background of the research method or approach that was used. For instance, 'research method' and 'research approach' were mentioned in only 114 and 18, respectively (121 for either term, 19% of the 653 papers), while the terms 'epistemology' and 'ontology' occurred in only 49 and 5 of them (51 for either, 8%).

Specific markers of quantitative approaches were 'questionnaire' (123 articles), 'factor analysis' (42), and 'structural equation model' (31); a total of 154 articles (24%) included at least one of these terms. Markers of qualitative approaches included 'interview' (177 articles), 'content knowledge' (94), 'constructivism' (34), 'focus group' (28), 'action research' (21), 'content analysis' (15), 'reflective practice' (9), 'phenomenology' or 'phenomenography' (9), and 'lesson study' (7); a total of 283 articles (43%) included at least one of these qualitative marker terms.

Overall, these results suggest that around a quarter of the papers (24%) gave evidence of quantitative approaches, and almost half (43%) gave evidence of qualitative approaches. A total of 358 papers (55%) made reference to at least one of these qualitative or quantitative markers, and 79 papers (12%) referred to both qualitative and quantitative markers—a possible indication of 'mixed methods' approaches, though the term 'mixed methods' itself occurred in only 21 of the papers (3%).

This summary presents a profile of statistics education research in which almost half of papers have no explicit discussion about research approach or research method, and very few of them make any reference to the philosophical aspects of such methods, though some of these may have done so implicitly or using some research approach that we did not include in our list of search terms. Using our marker terms, a larger group of studies seems to be making use of qualitative methods, and—maybe surprisingly—a smaller group utilizes standard statistical tools as their research approach. These conclusions may be limited by being based on the text of the collection of papers searched using the indicator terms that were identified as being feasible. Alternative approaches, such as careful reading of all 653 papers or a carefully selected random sample of these and other articles, might yield somewhat different conclusions.

3.4.2 The Object of Statistics Education Research (the **What***)*

In this section we will discuss the object of research in statistics education. We start with some background on the development from the early days to current times. Of course, what is meant by 'early days' is debatable, but as in Chap. 2, we start with the first ICOTS conference held in 1982, following previous discussions about statistics teaching at earlier ISI conferences. As noted in Chap. 2, the journal *Teaching Statistics* was started in 1979, *JSE* in 1993, and *SERJ* in 2000, though the newsletters of the *International Study Group for Research on Learning Probability and Statistics* go back to 1987, and the group itself even further. Here we focus on how research in statistics education has evolved from early discussions of pedagogical problems.

At the ICOTS 6 conference, Ottaviani's (2002) keynote address investigated the papers published in ICOTS 1–6 and showed on the basis of their titles alone how the focus of the ICOTS conferences had changed over time. The first two ICOTS conferences showed a focus on teachers and teacher training, predominantly in schools and sometimes using computers, and at tertiary level the cooperation between academic and practicing statisticians. ICOTS 3 placed students at the centre of interest, focusing on the materials and approaches to develop quantitative literacy at school and introductory statistics courses at university. ICOTS 4 and 5 continued this approach, increasing the integration of computer-based approaches, particularly for data analysis and developing statistical concepts. They also reported on the use of projects in developing students' experience and made explicit reference to students of applied disciplines such as engineering and economics. In her keynote address at ICOTS 6, Ottaviani identified for the first time a focus on research, not only research in teaching and learning statistics but also as 'research methods', a promising way of teaching students of different disciplines about statistics. The previous references to computers broadened into the application of technology, earlier notions of quantitative literacy became a broader focus on statistical literacy and thinking, and teacher training expanded to include development of statistics professionals generally.

At the same conference (ICOTS 6), Watson (2002) presented a summary of contemporary research in statistics education, utilizing a division into theoretical, qualitative, and quantitative studies. These she related to statistical thinking, statistical reasoning, and statistical literacy, based on her own work in these areas as part of the *Statistical Reasoning*, *Thinking*, *and Literacy* (SRTL) research forums (SRTL, 2017). As an example of theoretical research, she described Wild and Pfannkuch's

(1999) study on statistical thinking. As an example of qualitative research, she discussed the development of students' understanding of sampling derived from their responses to media articles (Watson & Moritz, 2000a). This project was extended to a survey of several thousand school students' ideas about sampling (Watson & Moritz, 2000b), providing an example of quantitative research.

Other sources of 'historical' information are previous reviews of statistics education research from participants in the field and for the field (e.g. Jolliffe, 2003) and also from/for other groups such as mathematics educators (e.g. Shaughnessy, Garfield, & Greer, 1996). The chapter by Garfield and Ben-Zvi (2008) also summarizes the historical background and then gives an overview of research questions addressed by contemporary researchers. The *American Statistical Association* report on research directions and priorities in statistics education (Pearl et al., 2012) mentioned earlier lists issues, research questions, and research priorities in six specific areas: cognitive outcomes, affective constructs, curriculum, teaching practice, teacher development, and technology, with a final section reviewing the range of currently available assessment instruments. Another recent review, focusing specifically on the use of technology at the school level, was carried out by Biehler, Ben-Zvi, Bakker, and Makar (2013) and may also take its place in the historical context.

Our examination using NVivo of 653 articles published between 2010 and 2014 also enabled us to explore the question of what is being currently investigated in the area of statistics education research. We used the following words or phrases, selected from the complete text of the collection of articles, as search terms for the investigation: curriculum, assessment, GAISE (*Guidelines for Assessment and Instruction in Statistics Education*), technology, statistical reasoning, statistical thinking, statistical literacy, attitudes towards statistics, conceptions (of statistics or some aspect of statistics), and probability.

The most common focus of statistical investigations in the collection of papers was 'curriculum', mentioned in 378 of the 653 articles. This may be due to the essential characteristics of statistics education research and particularly to the fact that the *small-r* research papers were most likely to be concerned with some practical aspect of teaching and learning statistics. The occurrences of 'curriculum' showed the term being used in a broad sense; related terms included 'assessment' (392 articles), 'technology' (369), and 'GAISE' (138). In total, 564 articles (86%) included one of these terms related to some aspect of curriculum.

Another very common aspect reported in these papers was 'statistical literacy' (250 articles), 'statistical thinking' (214), or 'statistical reasoning' (214). A total of 392 articles (60%) referred to at least one of these terms, sometimes collected together under the acronym SRTL (statistical reasoning, thinking, and literacy) and as such the subject of an active group of researchers. A total of 587 articles (90% of the papers) included at least one of the curriculum or SRTL terms.

Other topics in statistics education research were mentioned less frequently. They included the psychological aspects of 'attitudes towards statistics' (45 articles) and 'conceptions' of statistics or some aspects of statistics (140, but more accurately 70, since the other 70 made only one mention of the term, most often as part of a reference). For the most part, these results are based on students' ideas, but they sometimes include teachers' ideas (particularly in the overlapping situation where students are preparing themselves to be teachers). Although it was problematic to use 'probability' as a search term, due to its variety of uses, it did occur in over 400 papers but most often with its usual technical meaning rather than as an object of research. However, there were a number of papers focusing on aspects of teaching and learning probability or investigating students' ideas about probability.

This summary presents a picture of the objects of interest of statistics education research. They are predominantly focused on two main areas: the broad topic of curriculum, including assessment, technology, and teaching guidelines, as represented by GAISE and the broad notion of statistical literacy, sometimes described as 'statistical literacy, reasoning, and thinking' or 'statistical reasoning, thinking, and literacy' (SRTL). Less commonly, research investigates students' (and teachers') attitudes towards and conceptions of statistics or various aspects of statistics, and a number of studies focus specifically on aspects of probability. Our earlier comments about the limitations of the investigation apply here also.

3.4.3 The Participants in Statistics Education Research (the **Who***)*

The third aspect of research in statistics education pertains to the people involved in research about statistics education. There are, of course, a wide variety of people involved in such work, from individual (or teams of) teachers who are carrying out investigations on their own classes beyond the work for which they are paid (whom we refer to as *small-r* researchers) through to academics leading university-based research groups carrying out systematic investigations into some aspect of statistics education (*large-R* researchers). Also included are individuals or groups carrying out research on some aspect of statistics education but unconnected with the formal teaching of the discipline in schools and universities. The aim of this section is to give an indication of the various groupings of researchers in the world of statistics education research.

We are, of course, not the first to reflect on such issues. In 1995, when Garfield wrote her oft-cited paper 'How students learn statistics', there was little solid research within the field of statistics education itself; rather, most research came from studies or models in other areas of relevance, such as psychology or mathematics education. In 1998, and later in 2003, Jolliffe (who would later become the first co-editor, together with Batanero, of *SERJ*) reflected directly on the *who* question, as by that time research by statistics educators had begun to emerge. More recently, Garfield and Ben-Zvi's (2007) 'revisited' article (also the chapter 'Research on teaching and learning statistics' in Garfield & Ben-Zvi, 2008, pp. 21–43, and the discussion in Zieffler et al., 2011) gave a summary both in historical terms and of the then-current participants in statistics education research. During the time covered by these references, the field of statistics education has matured and expanded in several ways, and thus it is worthwhile revisiting questions about groups involved in research in or about statistics education.

There are various ways to consider such groupings, for instance, participants at the ICOTS conferences or regional and national conferences such as OzCOTS (Australian Conference on Teaching Statistics); authors, editors, and reviewers of a journal such as *SERJ* (particularly those working on a special issue); researchers belonging to a specific research programme, such as the team at the University of Minnesota headed by Garfield; researchers focusing on a particular aspect of statistics education research, such as SRTL (The International Collaboration for Research on Statistical Reasoning, Thinking, and Literacy, SRTL, 2017); or members of a particular statistics education initiative, such as CensusAtSchool.

One theoretical lens for such an endeavour is available from Wenger's (1998) notion of 'communities of practice'. According to this model, a community of practice consists of a group of people who share a domain of interest and who engage and interact with others in the group in their development of a shared repertoire to achieve their aims. Wenger (2013, p. 1) describes this succinctly on a website: 'Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly'. Statistics education researchers could be described as a single, large, and diverse community of practice, according to Wenger's ideas. Alternatively, it may be more useful to identify several professional (and social) groupings of people involved with various aspects of statistics education research, representing several interlocking communities of practice.

Some researchers may be members of more than one of these communities of practice, participating fully in quite different aspects of statistics education research, whereas others might be peripheral members of one or more such groups. For instance, Groth (2015) highlights the growing 'boundary interactions' between researchers in mathematics education and statistics education. The structure of such a group or groups may shape the topics and directions taken in research in the field. Individual statistics educators who are carrying out exploratory research on their own practice (*small-r* research) could be viewed as 'legitimate peripheral participants' (Lave & Wenger, 1991), newcomers to the community who are becoming acquainted with the tasks, vocabulary, and organizing principles of the community, on the way to becoming full participants. Such individuals may develop a coherent programme of research, working on their own or joining with other researchers with similar interests, and hence become more central participants in the community of practice.

Our analysis of the literature, helped by our familiarity with a variety of different projects and actors in the field, suggests that there are multiple ways to characterize the persons and groups doing research in or about statistics education. For simplicity, we focus below on three characterizations, (1) by disciplinary background of researchers, (2) by the institutional context of the research, and (3) by the geographical or cultural context of the research. However, our purpose is not to offer these as rigid groupings, since they are not mutually exclusive, and each of these labels is by itself multifaceted. Rather, we use these characterizations as part of our broader effort to sketch the broad diversity of *who* is doing research in our field, and to enrich our view of the uniqueness and scope of our field, and our understanding of how knowledge has evolved and may continue to develop in the field of statistics education.

3.4.3.1 Disciplinary Background of Researchers

It is useful to reflect on the disciplinary or professional background of the individuals involved in research in statistics education, in other words, the domain in which they have obtained their academic degrees, and the academic departments or organizations in which they work. Although there is no central database in which all relevant data are compiled, it is possible to get a sense for disciplinary or professional background by analysing contact information and biographical sketches of scholars who present research in journals such as *SERJ* or *JSE* or in conferences such as ICOTS.

When we speak to people outside the field of statistics education, they often express a belief that it is statisticians who do research on statistics education. However, even a cursory analysis of sources such as those outlined above suggests a very different picture. Typically, it involves people with backgrounds in statistics, psychology, and various fields within education, notably mathematics education, but also other STEM (science, technology, engineering, mathematics) fields such as computer education. In addition, there are professionals from other fields where statistics are being used and taught, such as biostatistics, biology, agriculture, business, medical sciences, and various social sciences, and also specialists from official statistics agencies (i.e. statistics producers) and other organizations who may be involved in initiatives with an educational component.

Although arguably incomplete, the list above is sufficient to suggest that the disciplinary background of those publishing research in or about statistics education is very broad and extends well beyond statisticians per se. This is of interest as researchers' disciplinary background may dictate what they count as credible evidence and may affect their choice of topics for research and the methods that are deemed appropriate to study them. For instance, the professional training of statisticians is normally focused on quantitative methods with less or little attention to the design of research instruments (i.e. questionnaires) and little or no attention to qualitative methods (Ograjenšek & Gal, 2016). As a result, statisticians may focus on research questions that are accessible to them or that can be examined within a positivistic paradigm, and much less so on research problems that require qualitative or mixed-methods methodologies, which are more common, for example, in mathematics education. We have seen earlier using our analysis of published articles that around a quarter of the papers gave evidence of quantitative approaches, and almost half gave evidence of qualitative approaches; this seems consistent with our observation that much research in statistics education is being carried out by workers outside the field of statistics.

3.4.3.2 The Institutional Context of Research

'Context' is a broad term and has various meanings when it comes to research. Here we focus on institutional aspects of the context. Firstly, researchers are usually defined by way of their work role, and this label primarily pertains to academics who are expected to produce research as an integral part of their professional work and to contribute to the accumulation of general knowledge via publication in peer-reviewed academic journals and conferences. Such *large-R* research usually addresses broad challenges not limited by time or place, is informed by broad theoretical models, and may aim to help theory-building or to have educational implications of a broad nature. However, in statistics education we also see many examples of *small-r* research, carried out by people, most often school or university teachers, whose motivation is mainly on local problems set in a particular context.

A typical and important example of *small-r* research is of teachers who may not define themselves as researchers yet are motivated by phenomena they encounter as part of their own practice to carry out an investigation on their own class, sometimes with locally designed instruments (such as a brief attitude survey or a test of specific cognitive tasks). Situations where teachers examine their own teaching in a systematic manner and share their conclusions with others via scholarly channels (e.g. conferences, peer-reviewed journals) have received increasing attention over the last three decades in the academic literature on the 'scholarship of teaching and learning' or 'scholarship research', pertaining to both tertiary/academic and school contexts (see Bennett & Dewar, 2012; Boyer, 1990).

Small-r research sometimes gets reported in professional conferences but we believe is not receiving wide representation in professional journals related to statistics education, despite its obvious importance to the actual improvement of practice and its potential to inform *large-R* research in a bottom-up fashion. This might be because of several reasons, such as teachers' lack of experience in writing for scholarly journals, because the customized nature of their instruments or research design makes it difficult to generalize in a way that is valued (or accepted) by referees in peer-reviewed journals, or because the researchers may be less familiar with the extant literature and hence may have difficulty explaining how their research fits into the broader body of existing knowledge.

This implies that the picture provided by papers in academic peer-reviewed journals is incomplete. What we know about key problems or processes in concept development and other issues that affect teaching and learning in statistics, about teachable moments, or about what works or does not work, and possible educational innovations, may be limited in several ways. Cumulative wisdom that may exist in the field and that may have been captured by localized *small-r* studies could be hiding under the radar of the *large-R* research community in statistics education. Countering this to some extent is the inclusion in our qualitative analysis of papers presented at conferences on statistics education, albeit at those with the highest profile (such as ICOTS); these papers certainly include examples of such *small-r* studies.

3.4.3.3 The Geographical or Cultural Context of the Research

Another possible aspect of the context of research is the particular country or region where research is being conducted and, linked with that, the particular academic culture from which the research arises. An examination of research in statistics education shows that a large proportion of research papers are written by people in English-speaking countries (primarily the USA, Australia, the UK, and New Zealand) or other European countries (e.g. the Netherlands, Spain, Belgium, France, etc.). Such wealthier countries differ from many less wealthy countries in terms of the quality of educational infrastructure, the availability of computer systems and advanced technology in the classroom, the quality of teacher education, and other organizational aspects such as the presence of national evaluation schemes or national curricula. These differences will have obvious effects on statistics education and thus on the statistics education research that is undertaken and the results obtained.

Certainly, the predominance of publications in English is a result of the scientific hegemony of the English-speaking world, in which research is published preferentially in English, the *lingua franca* of the academic world. Even researchers in countries with other languages will aim if possible to send their results to an English language journal or present them at an English-speaking conference; if they don't, their results will most often be relegated to 'local' (and thus, of limited interest) status. Of the sources that we used for our collection of articles for analysis, only *SERJ* gives researchers the opportunity to submit papers in another language (French or Spanish, but in fact during the 2010–2104 window, only one non-English paper was published, Bihan-Poudec, 2010, in French and with an English summary).

When Zieffler et al. (2011) analysed the output from the first 8 years of *SERJ*, the authors of the publications were drawn from only 15 countries. A special issue of *SERJ* at the end of 2014 (volume 13, number 2), entitled 'A Global View of Statistics Education Research', aimed specifically to broaden international representation (see North, Reston, Cordani, & Petocz, 2014), and currently 25 countries are represented. However, this is still a very small proportion of the world's countries and smaller than the representation at conferences such as ICOTS (at which around 50 countries may be represented).

In the 2014 *SERJ* special issue, there were eight papers published from Brazilian authors (perhaps a continuation of the interest generated by Brazil hosting ICOTS 7 in 2006). These papers gave an insight into a strong culture of statistics education research, based on a commitment to statistics education at all levels from pre-school to university. There was an obvious community of practice amongst Brazilian statistics education researchers, for the most part expressed in the Portuguese language. The *Associação Brasileira de Estatística* (ABE, *Brazilian Statistical Association*) is active with conferences and publications, and although none of them focuses only on statistics education research, they do include this as a topic of study. Several universities have active groups of researchers in the field, for instance, the University of São Paulo (USP) and the Federal University of Rio Grande (FURG), and research students are completing master's and doctoral degrees in aspects of statistics education. The *Sociedade Brasileira de Educação Matemática* (SBEM, *Brazilian Society of Mathematical Education*) has included since 2000 an active Working Group (GT12 *Ensino de Probabilidade e Estatística*, *Probability and Statistics Teaching*); a detailed description of its background and work is given by Cazorla, Kataoka, and da Silva (2010). The group has recently published, with the Pontifical Catholic University of São Paulo (PUC-SP), a special issue of *Educação Matemática Pesquisa* on research in statistical education (Coutinho & Samá, 2016). Despite all this, there was a significant effort involved in preparing and publishing eight papers from Brazilian authors in English, and without such effort these results are mostly unavailable to English-speaking researchers.

The French-speaking world also has a keen interest in statistics education research. The *Colloque Francophone Internationale sur l'Enseignement de la Statistique (CFIES)* has been held every 2 years (or so) since 2008 under the auspices of the *Société Française de Statistique*, although the papers presented do not seem to be available online. For Spanish-speaking researchers, particularly those in Latin America, the IASE publishes *Hipótesis Alternativa*, a bulletin that summarizes conferences, articles, and theses in Spanish (and sometimes Portuguese) to support the large community of Hispanic researchers in statistics education.

In other countries there is not only the problem of language but also that statistics education itself is only a very recent phenomenon; for instance, this very situation in Japan was described by Takemura (2013). With little teaching of statistics in schools, and no university department of statistics in the country, the absence of statistics education research was not surprising. However, the location of the 2018 ICOTS 10 conference in Kyoto will give a significant impetus to the development of statistics teaching and research in the country, and if it follows the trajectory of several other countries that have hosted ICOTS, the effects will be manifest for some time.

A related aspect of geography and culture has to do with social conditions or research traditions which predispose researchers to focus on specific issues or utilize specific methods. The 2014 special issue of *SERJ* provides some examples. Coming mostly from less-developed parts of the world, many papers showed a strong sense of social justice in the role of statistics education, or indeed education in general, in national development. The articles from South Africa demonstrated the use of statistics education to help in rebalancing social conditions after the years of apartheid, whereas many of the papers from Brazil showed elements of the 'critical pedagogy' of Paulo Freire (Freire & Clarke, 2000). In terms of research traditions, Spanish- and Portuguese-speaking authors seem to use a different palette of theory from that utilized in English-speaking countries (particularly the USA); several of the papers used the philosophical perspective of Piaget or the 'onto-semiotic' approach of Godino and Batanero (see Godino, Batanero, & Font, 2007).

The differences noted above may affect what is considered an educational problem and may cause researchers to take for granted certain contextual conditions in designing research or in discussing its relevance to other educators and teaching contexts. As a consequence, the results from research in some countries may have limited applicability to other countries, and judgements as to what is an important obstacle or problem for research may be affected by value judgements coloured by the geographical or cultural context.

This section has sketched in broad strokes various classifications of who is involved in research of relevance to statistics education. Coming back to the question of whether there is a single coherent community of practice, we believe that for now the answer is negative. The trends noted in this section regarding the great diversity in who is doing research could affect the potential impact of research in statistics education. Researchers tend to think that the purpose of research is to inform knowledge-building, but there are other views and critiques (e.g. Lester & Wiliam, 2002) of the actual contribution of educational research to practice and to educational policy both at the national level and at the local or institutional level (e.g. what curriculum to follow, what methods to adopt). We believe that there is much more to do before we can truly talk about an inclusive international community of practice in statistics education.

3.5 Discussion

In this chapter we have investigated overarching questions about statistics education research—the *how*, the *what*, and the *who*, and to some extent the *why*, of such research. Our investigation is both theoretical and empirical, the former looking at the underlying ideas of research approach and methodology, the latter based on an examination of a selected body of published research in the field. We face a wide range of research activities:

- In some cases, which we termed *small-r*, research in statistics education grows firstly from the curiosity of teachers of statistics working with their classes at all levels from pre-school to university. Some teachers turn this curiosity into investigations of their own classes and students or their own teaching practice, and this culminates in informal and contextualized research that is not designed for academic circulation.
- At the other end of a continuum are statistics educators who are involved in *large-R* research—purposeful, informed, methodical, and communicable investigations that are generally well beyond their own class or lecture rooms. They are joined in this endeavour by researchers who work in environments beyond the context of statistics education, maybe in some other discipline (such as mathematics) or for official statistics agencies.
- There are hybrid cases that mix aspects of *small-r* and *large-R*, such as evaluations of funded projects focused on statistics education or internal user surveys by official statistics agencies. Such efforts culminate in presentations or reports that may be quite formal yet are designed for internal circulation, or present conclusions with a local focus, with less a priori intention on deriving generalizable conclusions.

Regarding *how*, we have seen that methodologies for statistics education research can be described as positivistic or socially constructed, maybe incorporating a critical theoretic approach, and they can be broadly divided into quantitative and qualitative approaches and their various combinations. The quantitative, positivistic approaches, based on designed studies carried out with standard scientific approaches, are generally accepted as providing relatively rigorous conclusions—if and when they are carried out in a way that is consistent with the discipline of statistics itself. The qualitative approaches, based on case studies, interviews, ethnographies, content analyses, action research, and the like, have less status in scientific terms, and there are debates about the extent to which they provide an appropriate basis for pedagogical decisions (such as the US *No Child Left Behind Act* of 2001, see Shelley, 2005). Yet there are many situations where such qualitative investigations are more suited to the research goals, to actually finding out what is going on in the minds of participants in the educational process, and indeed they are in wide use and reported in many channels. It seems to us, however, that most published papers in the field avoid explicit discussion of the ontological aspects of their research methodology.

The question of *what* is being investigated in statistics education research has been addressed in several previous publications, but resulting descriptions need to keep up to date with shifting trends and emphases. Indeed, our empirical analysis of recent publications indicated the preponderance of studies that examine knowledge of curriculum topics and aspects of statistical reasoning, thinking, and literacy. Attitudes towards and conceptions of aspects of statistics were mentioned in a minority of articles.

In terms of *who*, we have utilized the notion of communities of practice to describe various interrelated groups of participants in statistics education research. We have concluded that the majority of research in the area is carried out by people who would not be described as practicing statisticians. For the most part, they are statistics educators in a wide variety of disciplinary fields and also people involved in the production and dissemination of statistical information. When statistics educators carry out *small-r* research based on their own context and practice, only part of this gets reported at professional conferences or journals, and there is a loss of potentially useful information. A minority of such educators are working in more formal research groups, usually at tertiary level, and publishing in academic journals. As most such publications are in English, results from researchers in countries with other languages, and sometimes other research traditions, are often difficult to include in the overall body of statistics education research.

Our reflection on the *how*, *what*, and *who* of research in statistics education has relied in part on an exploratory analysis of a collection of 653 articles published from 2010 to 2014. Such an analysis has not been previously reported in the literature; it has helped us to sketch in broad strokes some key features of research in statistics education. We have already outlined the rationale and some pros and cons of this analysis. Here we sketch some possible alternative directions for research that could expand our overall understanding of the methods and scope of research in or related to statistics education.

- The range of journals could be expanded to those that occasionally publish articles on statistics education research (such as *Mathematical Thinking and Learning* and other mathematics education journals). Likewise, the list of conferences could be extended to those that may include papers on statistics education research (such as the AERA, *American Educational Research Association* conferences). The reference range of years could be extended.
- A wider range of ideas and search terms could be used to capture methodologies and research on subtopics that are not easily identified by word frequencies or traditional word clouds.
- It may be possible to go beyond studies of core aspects of statistics education, to examine broader research that is not necessarily conceived by its authors as directly related to statistics education. Examples are studies on understanding of risk in domains such as financial or medical literacy, some of the research on human judgement and decision-making, research in mathematics education on understanding of ideas about proportionality and percents, research in science education about students' understanding of causality or research design, or research related to data literacy or design of dynamic visualizations.

These and other points imply that the search space for such a future bibliometric analysis or large-scale review of cumulative literature is nebulous and contestable. Although an overall picture such as the one we have given may have its uses, the best summary of specific areas of research in statistics education is given by people who have expert and comprehensive knowledge of each area: this is indeed what is given in the following chapters of this handbook.

3.6 Conclusions

The distinction between *small-r* and *large-R* research types, and the analysis of the diverse types and categories associated with the *how*, *what*, and *who* (and also the *why*) questions, seems important to us for several reasons. This analysis enables us to problematize what is considered 'research' in statistics education, what criteria are applied to judge the quality of research, and to whom and how such research is known and in turn how much it contributes to extant knowledge. The range of what qualifies as research in (or of relevance to) statistics education is broader than what gets published in leading journals and conferences in our field. Some research appears in publications of related disciplines, most commonly mathematics education research journals, seen in some circles (including, unfortunately, some funding bodies) as more prestigious than statistics education journals. Further, some types of research and potentially useful results 'fly under the radar' of the statistics education community, for instance, because of diverse perceptions regarding the acceptability of specific research questions or methodologies or due to the language of publication. In addition, self-posed intentions for generalization of the persons or groups involved in the research affect whether a research process is written up for publication and, if so, how it is written. These and other factors arguably contribute to loss of information and constrain the scope of what is known or what is accepted as ‘scientifically known’ in our field.

Another conclusion from the above analysis, as well as from the earlier exploration of the *who* question, is that there are several distinct communities of practice that generate research of relevance to statistics education, overlapping only occasionally. We believe that there is much potential to increase these overlaps and provide for better integration of efforts and cross-sharing of results between these communities of practice. This can certainly be aided by technological advances, such as quick and easy internet and video communication between researchers, and enhanced machine translation of articles and presentations from different languages, but it also depends on personal commitment from various members.

We advocate for and hope to see more aspects that improve knowledge-sharing in statistics education research, such as:

- Cultural inclusivity and tolerance in the peer review process.
- Mentoring of *small-r* researchers by those with more formal research skills.
- Reflective sessions at professional conferences that enable the sharing of practitioner research and programme evaluation efforts (i.e. seemingly *small-r* research) in ways that can emphasize their unique nature and contribution to cumulative knowledge and a discussion of their links with *large-R* research.
- ‘Research interpretation’ sessions that enable academic researchers and practitioners to examine *large-R* research from the point of view of teachers and teacher-trainers working in diverse settings. The goal in this case is to improve the contribution of the research to modes of thinking and to action plans of practitioners in the field. This would increase the likelihood that results from ‘academic’ research will actually be known and have an impact on educational decisions by those who are forming curricula and lesson plans, training teachers, or designing tools for assessment and monitoring of progress towards learning goals in statistics education.

Considering the future of research of relevance in statistics education, the chapter authors also offer individual views about challenges in this regard. The three authors have different disciplinary backgrounds (Peter in statistics and mathematics, Anna in music and higher education, Iddo in applied psychology); they have all been active in the field of statistics education research for some time (two decades or longer). Two of them (Iddo and Peter) have been editors of *SERJ* (and Anna has been editor of other journals). They come from different countries and have different first languages (Iddo from Israel, Peter from Hungary, and Anna from Australia). Between them they represent some of the diversity that we see in the field of statistics education research. Here is one final paragraph from each.

Iddo: One area for me is especially challenging: research on the development of statistical literacy of both adult populations and those who are still in formal education systems, tertiary, or school. Such research needs to cope with the multidisciplinary nature of statistical literacy, which encompasses not only knowledge bases in literacy and mathematics, statistics, and probability but also the many attitudinal and belief systems associated with statistical literacy. The development of statistical literacy, although important from a societal perspective, is facing many obstacles as it does not have a natural place in the curriculum or in the mind and schedules of many teachers (such as those who teach traditional introductory statistics at the college or high school levels). Hence research in this regard will also have to examine numerous institutional and applied aspects associated with curriculum change, teacher preparation, methods for evaluating the impact of professional development, or new technology tools, and more.

Anna: Of concern for the next few decades of statistics education research is its reach into other academic disciplines. Curiously, the examples provided in this chapter have a particular omission—very few creative and qualitative disciplines, such as music or design, are represented (one investigation of the area is reported in Gordon, Reid, & Petocz, 2014). Researchers from such disciplines find themselves on the periphery of the discussion and are often somewhat ill-prepared to use statistical ideas in their research because their discipline does not naturally include statistics in the early years of study. This is of particular importance when we consider that the *outcomes* of statistics education research are usually to *improve* an aspect of the learning environment that includes statistical thinking or practices, in order to *change* some aspect of contemporary life.

Peter: As an editor of the *Statistics Education Research Journal* for several years, I have had the opportunity of working with *large-R* researchers who are preparing the results of their investigations for publication and widespread dissemination. A truly wide coverage of aspects of statistics education research has been reported in that journal, and it is likely to continue as an up-to-date source of insight into such research. A particularly useful feature has been the special issues focusing on specific, and sometimes problematic, aspects of statistics education research. These have included research on reasoning about variability (2004), reasoning about distribution (2006), informal inference (2008), qualitative approaches in statistics education research (2010), attitudes towards statistics (2012), as well as a global view of statistics education research (2014) discussed earlier. A special issue on learning and teaching probability within statistics was recently completed (2016), and another on statistical literacy has just been published (2017). The topics of these special issues give a view of the key concerns for statistics education research over the past decade and more and will continue to act as signposts for future challenges.

Research of relevance to statistics education has come a long way over the last 30-some years and is a growing field that is developing its own distinct identity. Our analysis of questions pertaining to *how*, *what*, and *who* will most likely need to be re-examined in a few years in light of rapid changes in our field and the many contexts to which it is linked. The analysis (both the ontological and the empirical) suggests that it is not possible, nor wise, to set fixed or clear boundaries on what or who qualifies as 'research(ers) in statistics education'. The appearance of new needs and new contexts for teaching and learning statistics and probability (e.g. in areas such as health education, financial education, or civic education) will also require fresh thinking on accepted methods and the use of mixed-methods designs.

Amongst other things, expanding uses of technology are rapidly changing the landscape of statistics education. Increased use of specialized applets and dedicated software for teaching statistics raises new questions about teacher-student, student-student, and student-technology interactions and how to understand and improve them. The proliferation of megaclasses, MOOCs, and other virtual learning environments (such as training modules for the general public on websites of official statistics agencies or statistics providers such as Gapminder) raises new questions about the nature of teaching and learning processes. As Gal and Ograjenšek (2010) argue, these and related changes require the use of new research methodologies (e.g. netnography, text analytics, log analysis) and an expansion of the range of expertise or disciplinary background of those involved in research on statistics education.

References

ASA (American Statistical Association). (2007). *Using statistics effectively in mathematics education research (SMER Report)*. Retrieved from http://www.amstat.org/asa/files/pdfs/EDU-UsingStatisticsEffectivelyinMathEdResearch.pdf

Bennett, C. D., & Dewar, J. M. (2012). An overview of the scholarship of teaching and learning in mathematics. *PRIMUS (Problems, Resources, and Issues in Mathematics Undergraduate Studies), 22*(6), 458–473.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technologies for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel-Kreidt, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–690). New York: Springer Science+Business Media.

Bihan-Poudec, A. (2010). L'enseignement de la statistique: en premier lieu, l'apprenant (Statistical education: Focusing on the learner). *Journal of Statistics Education, 9*(2). Retrieved from http://iase-web.org/documents/SERJ/SERJ9(2)_Bihan-Poudec.pdf

Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. San Francisco, CA: Jossey-Bass.

Brantlinger, A. (2014). Critical mathematics discourse in a high school classroom: Examining patterns of student engagement and resistance. *Educational Studies in Mathematics, 85*(2), 201–220.

Cazorla, I. M., Kataoka, V. Y., & da Silva, C. B. (2010). Trajetória e perspectivas da educação estatística no Brasil: um olhar a partir do GT12 (The growth of and prospects for statistics education in Brazil: A view from the GT12). In C. E. Lopes, C. Q. S. Coutinho, & S. A. Almouloud (Eds.), *Estudos e Reflexões em Educação Estatística (Studies in and reflections on statistics education)*. Brazil: Mercado de Letras, Florianopolis.

Coutinho, C., & Samá, S. (2016). Editorial. *Educação Matemática Pesquisa, 18*(3), 1099–1108. Retrieved from https://revistas.pucsp.br/index.php/emp/issue/view/1658

Evans, M. (2010). Researcher practice: Embedding creative practice within doctoral research in industrial design. *Journal of Research Practice, 6*(2), 1–17. Retrieved from http://jrp.icaap.org/index.php/jrp/article/view/200/197

Freire, P., & Clarke, P. (2000). *Pedagogy of freedom: Ethics, democracy, and civic courage (Critical perspectives series: A book series dedicated to Paulo Freire)*. Lanham, MD: Rowman & Littlefield Publishers.

Gal, I. (2002a). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–25.

Gal, I. (2002b). Dispositional aspects of coping with interpretive numeracy tasks. *Literacy and Numeracy Studies, 11*(2), 47–61.
Gal, I., & Ograjenšek, I. (2010). Qualitative research in the service of understanding learners and users of statistics. *International Statistical Review, 78*(2), 287–298.
Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*(1), 25–34.
Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372–396.
Garfield, J., & Ben-Zvi, D. (2008). Research on teaching and learning statistics. In *Developing students' statistical reasoning: Connecting research and teaching practice* (pp. 21–43). New York: Springer Science+Business Media.
Godino, J., Batanero, C., & Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM: The International Journal on Mathematics Education, 39*(1/2), 127–135.
Gordon, S., Reid, A., & Petocz, P. (2014). Researchers' use of statistics in creative and qualitative disciplines. In H. McGillivray, M. Martin, & B. Phillips (Eds.), *Australian conference on teaching statistics—Topics from OZCOTS 2008–2012* (pp. 365–384). New York: Springer Science+Business Media.
Groth, R. (2015). Working at the boundaries of mathematics education and statistics education communities of practice. *Journal for Research in Mathematics Education, 46*(1), 4–16.
Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). London: Sage.
Hill, H. C., & Shih, J. C. (2009). Examining the quality of statistical mathematics education research. *Journal for Research in Mathematics Education, 40*(3), 241–250.
Hobden, S. (2014). When statistical literacy really matters: Understanding published information about the HIV/AIDS epidemic in South Africa. *Statistics Education Research Journal, 13*(2), 72–82. Retrieved from http://iase-web.org/Publications.php?p=SERJ
Jolliffe, F. (1998). What is research in statistics education. In *Statistical education—Expanding the network: Proceedings of the Fifth International Conference on Teaching of Statistics, ICOTS5* (pp. 801–806).
Jolliffe, F. (2003). Towards a data base of research in statistical education. *Statistics Education Research Journal, 2*(2), 47–58.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
Lester, F. K., & Wiliam, D. (2002). On the purpose of mathematics education research: Making productive contributions to policy and practice. In L. English (Ed.), *International handbook of research in mathematics education* (pp. 489–506). Mahwah, NJ: Lawrence Erlbaum Associates.
Makar, K., & Ben-Zvi, D. (2011). The role of context and evidence in informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1/2), 1–4.
Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.
Moore, D. S., & Cobb, G. W. (2000). Statistics and mathematics: Tension and cooperation. *American Mathematical Monthly, 107*(7), 615–630.
North, D., Reston, E., Cordani, L., & Petocz, P. (2014). A global view of statistics education research. *Statistics Education Research Journal, 13*(2). Retrieved from http://iase-web.org/documents/SERJ/SERJ13(2)_Editorial.pdf
Ograjenšek, I., & Gal, I. (2016). Enhancing statistics education by including qualitative research. *International Statistical Review, 84*(2), 165–178.
Ottaviani, M. (2002). 1982–2002: From the past to the future. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics, ICOTS 6*. IASE: Cape Town, South Africa. Retrieved from http://iase-web.org/documents/papers/icots6/07_ot.pdf
Pearl, D., Garfield, J., delMas, R., Groth, R., Kaplan, J. McGowan, H., & Lee, H.S. (2012). Connecting research to practice in a culture of assessment for introductory college-level statistics. Retrieved from http://www.causeweb.org/research/guidelines/ResearchReport_Dec_2012.pdf

Petocz, P., & Reid, A. (2010). On becoming a statistician: A qualitative view. *International Statistical Review, 78*(2), 271–286.

Popper, K. (1959). *The logic of scientific discovery*. London: Routledge.

Scotland, J. (2012). Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of scientific, interpretive, and critical research paradigms. *English Language Teaching, 5*(9), 9–16.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 205–238). New York: Springer.

Shelley, M. (2005, August). *Education research meets the gold standard: statistics, education, and research methods after "No Child Left Behind"*. Presented at the Joint Statistical Meetings, MN, USA. Retrieved from http://www.matstat.com/teach/Shelley.ppt

SRTL. (2017). Research forums. *International collaboration for research on statistical reasoning, thinking and literacy* [website]. Retrieved from srtl.info

Takemura, A. (2013). Activities of Japanese inter-university network for statistical education. *Joint IASE/IAOS satellite: Statistics education for progress*. Retrieved from http://iase-web.org/documents/papers/sat2013/IASE_IAOS_2013_Paper_K5_Takemura_ppt.pdf

Vita, A., & Kataoka, V. (2014). Blind students' learning of probability through the use of a tactile model. *Statistics Education Research Journal, 13*(2), 148–163. Retrieved from http://iase-web.org/documents/SERJ/SERJ13(2)_Vita.pdf

Watson, J. (2002). Doing research in statistics education: More than just data. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics, ICOTS 6*. IASE: Cape Town, South Africa. Retrieved from http://iase-web.org/documents/papers/icots6/03_wa.pdf

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3–46. Retrieved from http://iase-web.org/documents/SERJ/SERJ2(2)_Watson_Callingham.pdf

Watson, J. M., & Moritz, J. B. (2000a). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*(1), 44–70.

Watson, J. M., & Moritz, J. B. (2000b). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior, 19*(1), 109–136.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge, UK: Cambridge University Press.

Wenger, E. (2013). *Communities of practice: A brief introduction*. Retrieved from http://wenger-trayner.com/wp-content/uploads/2013/10/06-Brief-introduction-to-communities-of-practice.pdf (or just refer to the website at http://wenger-trayner.cp/theory/).

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–265.

Zieffler, A., Garfield, J., delMas, R., Le, L., Isaak, R., Bjornsdottir, A., et al. (2011). Publishing in SERJ: An analysis of papers from 2002–2009. *Statistics Education Research Journal, 10*(2), 5–26. Retrieved from http://iase-web.org/documents/SERJ/SERJ10(2)_Zieffler.pdf

Part II
Major Contributions of Statistics Education Research

Maxine Pfannkuch and Robert delMas

Part I of the Handbook presents an overview of Statistics and Statistics Education and then situates Statistics Education Research within both of those perspectives. This part of the Handbook, Part II, aims to provide a discussion of major contributions of statistics education research related to learning and understanding statistics. Rather than having each chapter focus on a specific statistical topic (e.g., hypothesis testing), seven themes were chosen that highlight not only the research conducted within statistics education but also the gaps in the research knowledge base. The seven themes covered in Part II are the practice of statistics, research on data, research on uncertainty, introducing children to modeling variability, learning about statistical inference, statistics learning trajectories, and research on statistics teachers' cognitive and affective characteristics. Each chapter summarizes foundational and current work to illustrate contrasting perspectives, directions, and progress in our understanding of how students learn key statistical concepts. The authors of each chapter also identify promising methodologies and questions that need to be addressed and explored to further our understanding through future statistics education research. This sets the stage for Part III of the Handbook where promising new approaches and perspectives on both the learning and teaching of statistics and methodologies for future statistics education research are discussed.

The first theme considers the enterprise of statistics as a whole, the investigative cycle of inquiry and statisticians' practice, and how students might be enculturated into statistical thinking and practice. In Chap. 4, Jane Watson, Noleine Fitzallen, Jill Fielding-Wells, and Sandra Madden first describe several different frameworks that have been proposed by national guidelines and researchers for the practice of statistics, highlighting similarities and differences across the frameworks. This is followed by summaries and descriptions of a variety of research on stages of the statistical investigative cycle that are common across the different frameworks (e.g., problem posing, data production, visual representations, technology, data analysis, statistical inference), as well as studies of students engaged in the entire investigative cycle. The authors point out areas that have not been researched (e.g., students understanding and ability to carry out the structuring, cleaning, and manipulation of data) as well as numerous areas where future research is needed to corroborate and extend research findings, especially those from small-scale studies. Several *Big Ideas* of statistical thinking and practice (e.g., Data, Center, Variability, Sampling, Statistical Models) that permeate and unify the different stages of a statistical investigation are discussed. Chapter 4 ends with a discussion of statistical literacy, raising the question of whether students can develop adequate statistical literacy without having experienced the practice of statistics. The thorough coverage of research will provide the reader with a rich understanding of effective practices for promoting students' understanding of the practice of statistics, as well as areas where our knowledge of how best to teach statistics is wanting.

The next four chapters cover themes that we regard as key components underpinning statistical inquiry. The focus of the second theme is on the role and use of data in statistical inquiry, summarizing research on students reasoning with and about data. In Chap. 5, Rolf Biehler, Daniel Frischemeier, Chris Reading, and Mike Shaughnessy first explore what it means to reason about data from different perspectives such as national curricular documents and association guidelines, as well as research frameworks. The discussion about reasoning is followed by sections organized around four statistical concepts related to data: variability and variation, distribution, comparing groups, and association. Each section provides a thorough overview of research methodologies and frameworks that have been used to investigate students' reasoning at different ages and for different groups (e.g., preservice teachers), and our current understanding of how reasoning about data in each area develops. Several of the sections highlight the role of technology in the development and study of students' reasoning and thinking about data. The chapter ends with a summary of the findings and a discussion of the pros and cons of the research methodologies that have been used to study reasoning about data, with recommendations for methods that can expand our understanding.

The third theme looks at probability and uncertainty. Dave Pratt and Sibel Kazak look at the research on the teaching and learning of uncertainty in Chap. 6. Three primary issues from the research literature are emphasized: new theoretical perspectives on heuristics and biases in reasoning about uncertainty resulting from critiques of research in this area; the role of conceptual and experiential engagement in the conceptual development of reasoning about uncertainty, including the role of technology; adopting a modeling perspective to teaching and learning about probability. The role of the teacher in shaping the learning environment in various critical ways emerges as a key finding. The chapter concludes by identifying challenges to teaching uncertainty, needed areas (e.g., subjective probability) and promising directions (e.g., exploratory research on the role of modeling) for future research, and the need for carefully designed experiments to test hypotheses that are apparent from more established studies.

The fourth theme looks at the use of models and modeling in statistics instruction. Chapter 7 takes a more focused look at the potential role of what Richard Lehrer and Lyn English refer to as "inducting elementary aged children into the statistical practice of modeling variability." This chapter intersects with topics covered in previous chapters such as data, distribution, and variability, as well as the role of technology. There is a direct tie back to Chap. 6 through an emphasis on the use of models and modeling approaches. Summaries of conceptual frameworks for models, representations, and data modeling are presented. This is followed by a thorough review of the research literature from which emerges a trajectory of modeling experiences for developing elementary aged children's understanding of statistical concepts, including statistical inference. Suggestions for future research on the role of modeling and modeling technologies to expand students' understanding of statistical practice are discussed.

The fifth theme examines what we currently know about teaching students to understand statistical inference. In Chap. 8, Katie Makar and Andee Rubin review research on the learning of statistical inference, focusing in particular on informal statistical inference at the school level. The chapter begins by arguing for the importance of understanding statistical inference and the opportunity inference provides for unifying the learning of statistics. Research on the challenges students encounter in learning statistical inference is summarized, with emphasis on strategies that capitalize on technology. The chapter then turns to the research on informal statistical inference that has emerged over the past decade. Classroom research on approaches to developing informal statistical inference at primary, secondary, and tertiary levels and the impact of these approaches on understanding statistical inference are reviewed. The chapter ends by outlining gaps in research on statistical inference, suggestions for future research with emphasis on statistical modeling and big data, and the potential for informal inference approaches to reinvent the teaching and learning of statistics.

As illustrated in some of the previous chapters, the field of statistics education research has matured to the point where there is a better understanding of how some statistical concepts develop. When further considering the main contributions in the statistics education field, the development of learning trajectories to study statistical understanding and the role of the teacher in students' learning are prominent as major themes in current research. Exploring the sixth theme of Part II, Pip Arnold, Jere Confrey, Ryan Seth Jones, Hollylynne Lee, and Maxine Pfannkuch discuss in Chap. 9 how learning trajectories derived from knowledge and models of conceptual development can inform both teaching and research in statistics education. According to the authors, a learning trajectory is derived from a web of information on theories about statistics teaching and learning, knowledge of learning in the statistics context, and knowledge of statistics activities and representations. The authors outline the characteristics of learning trajectories and exemplify how learning trajectories have been used in research using three case studies: sixth-grade students' exploration of ways to represent and measure variability in data; preparing ninth-grade students to make a judgment when comparing two groups; teachers' conceptual development of repeated sampling in an inference context. Commonalities and differences across the learning trajectories in the three case studies, the potential of research based on learning trajectories to impact curriculum and classroom practice, current limitations and issues associated with this type of research (e.g., scalability and lack of pathways across grade levels), and implications for future research are discussed.

The seventh and final theme in Part II turns attention specifically to teachers of statistics. Randall Groth and Maria Meletiou-Mavrotheris review the literature on statistics teachers' cognitive and affective characteristics in order to understand the role they play in developing students' understanding of statistics in Chap. 10. The chapter first outlines and defines several frameworks for characterizing cognitive constructs (e.g., subject matter knowledge, pedagogical content knowledge, technological pedagogical content knowledge) and affective constructs (beliefs and attitudes). This is followed by an overview of different methods that have been used to assess statistics teachers' cognitive and affective characteristics (written assessments, interviews, observational studies). After a thorough review of frameworks and assessment methods, Groth and Meletiou-Mavrotheris summarize the research on teachers' understanding of key conceptual areas explored in previous chapters of this handbook (e.g., data, distribution, variability, association, uncertainty, inference), pointing out the similarities and differences between students' and teachers' understanding and attitudes in each of these areas, and identifying promising directions for teacher development. The chapter concludes with an exploration of promising methods for teacher education and development that promote more effective methods of teaching statistics.

As the editors of Part II of this handbook, we have had the privilege of shepherding each chapter through reviews and revisions. We are impressed with the breadth of the research that is covered by these seven chapters, as well as the quality of the writing and insights that the respective groups of authors have produced. Our understanding of statistics education and the conduct of statistics education research has been enriched by the process, and we hope that you will find your own understanding expanded as you engage with each chapter.

Chapter 4
The Practice of Statistics

Jane Watson, Noleine Fitzallen, Jill Fielding-Wells, and Sandra Madden

Abstract This chapter presents an overview of the Practice of Statistics focusing mainly on research at the school level. After introducing several frameworks for the practice, research is summarized in relation to posing and refining statistical questions for investigation, to planning for and collecting appropriate data, to analyzing data through visual representations, to analyzing data by summarizing them with specific measures, and to making decisions acknowledging uncertainty. The importance of combining these stages through complete investigations is then stressed both in terms of student learning and of the needs of teachers for implementation. The need for occasional backtracking is also acknowledged, and more research in relation to complete investigations is seen as a priority. Having considered the Practice of Statistics as an active engagement by learners, the chapter reviews presentations of the Big Ideas underlying the practice, with a call for research linking classroom investigations with the fundamental understanding of the Big Ideas. The chapter ends with a consideration of the place of statistical literacy in relation to the Practice of Statistics and the question of the responsibility of the school curriculum to provide understanding and proficiency in both.

Keywords Big Ideas • Complete investigations • Investigation frameworks • Posing questions • Statistics at school • Statistical literacy • Visual representations

J. Watson (✉) • N. Fitzallen
Faculty of Education, University of Tasmania, Hobart, TAS, Australia
e-mail: Jane.Watson@utas.edu.au; Noleine.Fitzallen@utas.edu.au

J. Fielding-Wells
School of Education, The University of Queensland, Brisbane, QLD, Australia
e-mail: j.wells2@uq.edu.au

S. Madden
Teacher Education and Curriculum Studies Department, University of Massachusetts Amherst, Amherst, MA, USA
e-mail: smadden@umass.edu

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_4

4.1 Introduction

The *Practice of Statistics* is the title of the well-known introductory college textbook of Moore and McCabe (1989), which is now in its eighth edition (Moore, McCabe, & Craig, 2014). In the first edition, Moore and McCabe said it was their "intent to introduce readers to statistics as it is used in practice. Statistics in practice is concerned with gaining understanding from data; it is focused on problem-solving" (p. xi). As a text written for the transition between the secondary and tertiary levels of education, this simple description suits all levels of education. The Practice of Statistics is hence the title and main focus of this chapter. It is a fortunate coincidence that the same year, the US National Council of Teachers of Mathematics (NCTM) published its *Curriculum and Evaluation Standards for School Mathematics* (1989), which introduced statistics across all years of schooling from kindergarten, again from a problem-solving perspective. Significant for educators and researchers was the final paragraph of Standard 10 for Grades 9–12:

> Statistical data, summaries, and inferences appear more frequently in the work and everyday lives of people than any other form of mathematical analysis. It is therefore essential that all high school graduates acquire, at the appropriate level, the capabilities identified in this standard. This expectation will require that statistics be given a more prominent position in the high school curriculum. (p. 170)

Although the *Standards* and Moore and McCabe (1989) were about curriculum and content, they opened the door to an era of research in statistics education spanning the entire range of education starting when children begin their schooling.

The Practice of Statistics as carried out by most professional statisticians is based on theoretical foundations and complex procedures for handling problems with data that are not accessible to school students. The thesis of this chapter, built upon the research reported, is that the intuitions that underpin the Practice of Statistics for students can be developed from the early years of schooling, preparing students for higher level courses and/or life as statistically literate citizens. Throughout Section II of the handbook, the focus is on the research associated with the manner in which this goal can be accomplished. In particular, this chapter examines the research on students' enculturation into the Practice of Statistics at the school level before encountering a course based on Moore et al. (2014).

This chapter initially considers several frameworks that researchers are using associated with the Practice of Statistics, before looking in more detail into research at the school level based on five stages: problem posing, planning for and collecting data, data analysis via visual representation, data analysis via reducing data, and drawing conclusions. Included at the end of the section is consideration of research investigating students carrying out the entire process in one activity. Recognizing the complexity of carrying out the Practice of Statistics, the following section focuses on suggestions for the Big Ideas (or fundamental concepts) that underpin the Practice of Statistics. The implications of the Practice of Statistics for statistical literacy more generally are considered with the research in that area in the final section.

4.2 Frameworks for the Practice of Statistics

Over the years, various frameworks have been suggested to describe the Practice of Statistics. At the school level in 1980, Holmes was instrumental in suggesting a statistics curriculum for schools in England and Wales based on five components: data collecting, data tabulation and representation, data reduction, probability, and interpretation and inference. In the United States following their professional learning work with teachers in support of the NCTM (1989) *Standards*, Bright and Friel (1998) produced a complex concept map surrounding the four main steps of the "process of statistical investigation." These steps were pose the question, collect the data, analyze the data, and interpret the results. From a different starting point, using analysis of the work of their university statistics colleagues, Wild and Pfannkuch (1999) suggested five stages in an investigative cycle when carrying out a statistical investigation: Problem, Plan, Data, Analysis, and Conclusion (PPDAC) (cf. Chap. 1). From these three perspectives, the Practice of Statistics involves carrying out a complete investigation. Implicit in all three frameworks is variation, which is made explicit in the four-step framework provided in the *Guidelines for Assessment and Instruction in Statistics Education* (*GAISE*) *Report* (Franklin et al., 2007) for the school level. For every step—formulate questions, collect data, analyze data, and interpret results—the role of variability is emphasized: *anticipating* variability in writing the question, *acknowledging* variability in designing the data collection method, *accounting for* variability in using distributions, and *allowing for* variability in interpreting the results. Although none of the other frameworks preclude repeating the process as a cycle, the title used by Wild and Pfannkuch (1999), Investigative Cycle, emphasizes the reality that professional statisticians face when the conclusion of one investigation raises questions leading to another. Those questions are often dependent on the context of the investigation.

The need for the recognition of a context within which to practice statistics is often taken for granted by researchers. Unfortunately, it is also sometimes forgotten by textbook writers who only present algorithms for procedures such as finding the mean. Rao (1975) recognized this when he wrote:

> Statistics ceases to have meaning if it is not related to any practical problem. There is nothing like a purely statistical problem which statistics purports to solve. The subject in which a decision is made is not statistics. It is botany or ecology or geology and so on. (p. 152)

Wild and Pfannkuch (1999) include context in one of the other dimensions of their framework, types of thinking. The need for context presents statistics education researchers with dilemmas in terms of what knowledge students are expected to have of the chosen context, of how much freedom students are given in choosing a context, and of how much time is allocated to carry out investigations.

With these frameworks suggested as a basis for the Practice of Statistics, the tools employed across the steps determine the type of conclusion that can be drawn. The framework of Wild and Pfannkuch developed at the tertiary level is likely to use more sophisticated tools and theory to reach a formal inference. This led Makar and Rubin (2009) to introduce an informal inference for younger students, accepting less sophisticated tools for analysis of data and acknowledging uncertainty in using evidence from data in a sample to reach a generalization about a population. The question that now arises is whether this is a description of the final stage of interpreting the results of an investigation or a move to a new framework (e.g., Makar, Bakker, & Ben-Zvi, 2011). Chapter 8 refines informal inference further including recognition of context and of the importance of aggregates, suggesting another alternative framework to consider.

Providing a visualization of the Practice of Statistics including all of the relevant facets would be very complex (cf. Bright & Friel, 1998; Watson, 2006). An attempt to portray the Practice of Statistics for the school classroom is shown in Fig. 4.1 (Watson, 2016), emphasizing the acknowledgment of uncertainty (rather than proof as is the norm in mathematics). The point at which students enter the investigation may vary, but the goal is that they reach meaningful and satisfying conclusions. The potential for research at every stage and overall is nearly endless. Progress in this arena is the focus of the next section.

4.3 The Practice in Action

Given the slight variations in the descriptions of the Practice of Statistics described in the previous section, the following five subsections address research related to the *GAISE* framework (Franklin et al., 2007), with the splitting of data analysis into two parts recognizing the importance of school students having explicit experience with visual representation as well as with the summarizing of data with statistics. The subsections also correspond to Fig. 4.1 with problem posing inferred by the statistical question, the level of uncertainty and informal inference combined, and variation underpinning the entire process. The section finishes with considering the importance of completing an entire investigation.

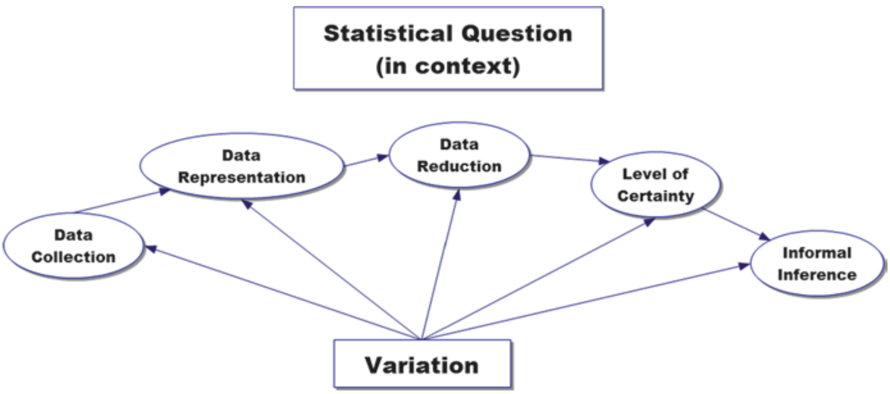


Fig. 4.1 A framework for statistical investigation (Watson, 2016)

4.3.1 Problem Posing: Asking, Understanding, and Refining the Statistical Question

In terms of the educational research associated with the Practice of Statistics, formulating statistical questions, that is, asking and understanding the question, has received little attention. There are several reasons for this. The statistical procedures that tend to be taught in school are those that are applied at later stages of an investigation, such as producing representations or finding measures of center and spread. These have often been the focus of the instruction and the research, and hence, the question and its context are set up for the learner. Starting with a preset question can also be seen as saving time as "think of a question" can set a very nebulous and difficult task for beginning students, even if the context is set. The *GAISE* description (Franklin et al., 2007) proposes a partial reason for the difficulty of formulating questions, where "anticipating variation" is the key feature and "requires an understanding of the difference between a question that anticipates a deterministic answer and a question that anticipates an answer based on data that vary" (p. 11).

In statistics education where problem posing sets the stage for an entire statistical investigation, the question arises as to where the starting point is in terms of a learning sequence. Is the starting point the context? Is it the context combined with a series of appropriate and inappropriate questions? Is it the context combined with a series of appropriate questions appreciating the type of data and variability involved? In other words, where are students expected to start and how much instructional help are they given before they start? Arnold (2008) made an initial distinction in this regard related to the purpose of the questions posed: the *investigative* question is posed to interrogate the data, and the *survey* question is posed to obtain the actual data. These are considered next.

Allmond and Makar (2010) focused only on the investigative, question-posing part of a statistical investigation in a study that included an eight-lesson unit on creating investigative questions. These 9-year-old Australian students considered characteristics of investigative questions, sorted questions into two types using the criteria described by *GAISE*, envisaged the data required to address the question, refined their questions, and considered the importance of the potential solution. This study provided an example of backtracking (cf. Konold & Higgins, 2003). Adapting the work of Arnold (2008), the study then compared the levels of questions created on pre- and post-tasks using a seven-point hierarchy: irrelevant, non-mathematical, non-investigative, closed, potentially investigative, investigative, and inquiry (acknowledging uncertainty). Results showed much greater improvement of the students who experienced the lessons over those in control classes that had no lessons on problem posing. Here, the framework for preparation for the final task was exceedingly explicit. The learning environment could be considered semi-structured in that students created their questions for a specific context set by the teacher.

A similar study based on investigations in science was carried out with Grade 6 Singapore students by Chin and Kayalvizhi (2002). They included instruction between a pre-problem-posing session and a post-problem-posing session. The instruction included examples of three experimental designs related to botany. As with Allmond and Makar (2010), they found students able to move from non-investigative to investigative questions after seeing the examples. For this study, the choice of context was free, and mostly students chose questions from a science context. In summarizing their research and that of others, Chin and Kayalvizhi suggested a nine-element typology of investigative questions. These covered comparison, cause and effect, prediction, design and make, exploratory, descriptive, pattern seeking, problem-solving, and validation of a mental model (p. 278). Although requiring some amplification, these types of investigation cover many kinds of experimental processes that can be the basis of problems posed. The *GAISE* report (Franklin et al., 2007) gives a three-stage developmental sequence of investigations across Levels A, B, and C of the *Guidelines* related to science starting with performing a simple experiment, then carrying out a comparative experiment, and culminating with an experimental design with random assignment (p. 17). No other published research was found providing school students with an open-ended statistical task similar to the research of Allmond and Makar or Chin and Kayalvizhi. Despite well-supported suggestions, for example, by Finzer and Parvate (2008) and Hammerman (2009), for introducing large multivariate data sets in the high school years in meaningful contexts and asking students (or teachers) to pose questions, perhaps about relationships, such activities were not discovered as subjects of research reports.

Another free environment approach was used by Zakaria and Salleh (2012) in exploring teachers' ability to pose questions based on a raw data set of 20 numbers. A total of 175 Malaysian mathematics teachers were given 20 min to imagine contexts and pose as many statistical questions as possible. The teachers posed 365 questions, 74% of which were considered as appropriate for a statistical investigation. Central tendency was the most common single topic chosen, and some problems included two or three topics. Few of the questions asked for a conclusion. Among the contexts suggested were football scores, books read by a class, and number of children in families.

After introductory sessions with two small groups of Canadian Grade 7 students, Lavigne and Lajoie's (2007) starting point was introducing experimental and survey designs and providing a "library of exemplars." The groups then devised their own investigative question, assisted by four prompts related to clarity of the question, the variable used, categories if used, and the population. Both groups decided on a survey design, posing one question to their classmates for data collection. For this study, it could be argued that the survey question and the investigative question were the same.

A different starting point was used by English and Watson (2015) who set the context for problem posing as investigating the school playground for four classes of Australian Grade 4 students. The students were to develop survey questions to find out their peers' thoughts on the school playing area. There was initial discussion with the class on their favorite neighborhood playground and reasons for it being enjoyable. Students then worked in groups of four, each student posing a question with four potential multiple-choice responses, for example: "What is your favorite thing to play in the playground? (a) Tag. (b) Offground tag. (c) Hide and Seek. (d) Ghost hunter." These were discussed and refined within the groups, with one question from each group being asked to the class, and then analyzed by the original groups. The problem posing of English and Watson was much more specific than that of Lavigne and Lajoie (2007), and there were no further prompts for how the questions should be monitored during the activity. For these Grade 4 students who were just beginning their experience with the Practice of Statistics, there was no discussion of a population (as done by Lavigne & Lajoie) wider than their school.

As other stages of frameworks for the Practice of Statistics become saturated with research studies, more attention can be directed at this initial stage, even considering how much and what information students can access. Three other aspects of research on problem posing offer promise for future research on the Practice of Statistics. One is the point at which the student "enters the game": how much information, if any, on the overall context is presented before the question(s) is/are posed. Second is how much general modeling and practice are provided as scaffolding before the task is set and during refining of the question. As little has been reported on the refining process while posing questions, this may be a third source of motivators for research.

4.3.2 Planning for and Collecting Data: A Focus on Samples and Sampling

The Practice of Statistics often involves collecting sample data from one or more population(s) and then making inferences about the population(s) from the findings gleaned from the sample(s). Watson (2006) suggests that "the purpose of a sample is to show the variation in a population so it can be characterized and summarized" (p. 28). Key to being able to make inferences about a population is the selection of the sample, the sample size, and acknowledgment of sampling variability (Pfannkuch, Arnold, & Wild, 2015). Issues related to these key ideas are addressed in the planning phase of an investigation.

When planning an investigation, students often rely on their intuitions about representativeness of samples developed from previous experiences that may be based on notions of fairness (Jacobs, 1999; Meletiou-Mavrotheris & Paparistodemou, 2015; Rubin, Bruce, & Tenney, 1990). Jacobs studied Grade 4 and Grade 5 students' understanding of sampling prior to instruction by presenting them with a variety of sampling methods, such as conducting a raffle. About half of the students made decisions on the basis of the representativeness of the samples and their potential to avoid biased results, whereas most of the rest based decisions on other issues, such as those related to practicality or notions of fairness. An exploratory study conducted by Rubin et al. included interviews of 12 senior secondary students who had not taken any statistics courses. Similar to Jacobs's study, many students focused on notions of fairness rather than considering the likelihood of a sample being representative.

Given opportunities to develop understanding of sampling concepts, students have been seen to show marked improvements (Osana, Leath, & Thompson, 2004; Watson & Kelly, 2005). In a longitudinal study based on classroom lessons and using survey questions similar to those in Jacobs's (1999) study, Watson and Kelly found students' notions about samples improved in the post-instruction results as they progressed from Grade 3 to Grade 5. Osana and colleagues also reported improvement in students' ability to use survey data rather than their own personal opinions and experiences. In a study across Grades 3, 6, and 9, Watson and Moritz (2000a) observed a six-step "hierarchy of increasing sophistication concerning sample size, selection, and resulting representativeness" (p. 63) across the grades, which they situated within a statistical literacy hierarchy.

More recently, research on sample and sampling variation has focused on students' reasoning from samples when making informal inferences. This has shifted the research emphasis from a focus on sample size to sampling variability (e.g., Pfannkuch et al., 2015). To some extent, this has been facilitated by the availability of technological tools that give students easy access to simulation tools and multiple visual representations of data. Gil and Ben-Zvi (2010) found Grade 6 students' ideas about random sampling were challenged when generating multiple random samples from a population using *TinkerPlots: Dynamic Data Exploration* (Konold & Miller, 2015). The students were concerned that the different random samples showed different, sometimes contradicting, results. This appeared to undermine the students' confidence in the results when making informal inferences. Conversely, Saldanha and McAllister (2014) used that sample-to-sample variability purposefully to have students assess their confidence in informal inferences made about a population during an intervention that involved the exploration of the variability of samples of increasing size from populations of lengths of genetically modified and normal fish. A difficulty experienced by students was keeping mental track of the multi-tiered resampling, which resulted in confusing differences in median length with actual median lengths. Manor and Ben-Zvi (2015) developed this area of research further, introducing an "integrated modeling approach" to explore sampling distributions with two Grade 7 students.

Most of the research on students' understanding of samples and sampling has not been set within the context of the students planning investigations. One exception is the study conducted by Meletiou-Mavrotheris and Paparistodemou (2015) who provided Grade 6 students the opportunity to make informal data-based inferences from self-generated statistical inquiries. The researchers found the intervention supported the students to "further appreciate the principles underlying sampling, and particularly the need for adequately large sample size, and for a random-based sampling procedure" (p. 401). The class showed improved reasoning about the meaning and role of sample, role of sample size, potential sources of bias, representativeness of samples for population attributes, and drawing conclusions from multiple surveys. It also illustrated the difficulties students have transferring understanding of samples and sampling gained from examples provided through initial classroom instruction to a real-life investigation, particularly when the context of the initial investigation is very familiar to the students. The shift students make from working in an instructional learning environment designed to lay the foundations of samples and sampling to the enactment of those key concepts in a student-driven statistical inquiry is worthy of further investigation.

To understand further the way in which students make decisions about samples and sampling within the planning phase of the Practice of Statistics, further research studies need to extend beyond providing scenarios on which students comment, which is demonstrated in Jacobs's (1999) study. Studies that allow students to work through inquiries, as in the Meletiou-Mavrotheris and Paparistodemou (2015) study, should provide learning experiences that mirror the actual Practice of Statistics. The special issue of *Educational Studies in Mathematics*, in which the paper of Meletiou-Mavrotheris and Paparistodemou was included, was titled "Statistical reasoning: Learning to reason from samples." As pointed out by Ben-Zvi, Bakker, and Makar (2015) in the lead article, the implications of initial sampling influence the entire following investigation, and this is the focus of much of the research reported in the special issue. The question can be raised, however, about whether sufficient research has been carried out on the fundamental ideas associated with sampling itself and subsequent application within an inquiry. In addition, research studies themselves require larger sample sizes to confirm the findings reported from the small-scale studies cited in this section. With the exception of studies by Watson and her colleagues (2000a, 2005) and Jacobs (1999), studies on students' understanding of sample and sampling report findings from very small groups of students.

4.3.3 Data Analysis: Devising and Presenting Visual Representations

One of the issues that practicing statisticians face is that of cleaning data once collected. At the school level, students are generally provided with data that are well-behaved. Despite the topic being mentioned in the New Zealand curriculum at Level 5 (Ministry of Education, 2009), no research was found that dealt specifically with this issue. The Australian Bureau of Statistics (2011) provides raw data from its *CensusAtSchool* site that have not been cleaned, which could provide a basis for research into students' ability to clean data, as well as how to analyze them after being cleaned. Another issue is structuring data, one that statisticians are recognizing as a problem area that they need to address. How students record and organize data is an area of research that is only just emerging (English, 2012; Konold, Finzer, Kreetong, & Gaston, 2014).

Perhaps, the most significant changes in the Practice of Statistics and its teaching in the last decade have been brought about by the availability of technological tools for generating, manipulating, and representing data. "Every piece of statistical information needs a representation—that is, a form. Some forms tend to cloud minds, while others foster insight" (Gigerenzer & Edwards, 2003, p. 258). When analyzing data, useful representations are often graphical or visual and analysis often benefits from viewing data in different distributional forms. Visual representations are used to tell "the story" of the data, to make meaning of statistical and contextual relationships, and to communicate patterns found in data (Monk, 2003). Wild and Pfannkuch (1999) coined the term "transnumeration" to describe this process as one of their types of statistical thinking. In the spirit of the analysis phase of the Practice of Statistics, the focus should be on creating visualizations that "do the best job of telling the story sharply and fairly" (Konold & Higgins, 2003, p. 202). As Gigerenzer and Edwards demonstrate, the use of confusing representations can lead to a variety of poor decision-making. In contrast to this, Shaughnessy and Pfannkuch (2002) and Chick, Pfannkuch, and Watson (2005) provide examples of contexts where transnumeration is a pathway to clarity and understanding.

As data sets grow larger and technology continues to evolve, the Practice of Statistics in schools has the potential to include the use of technological tools to support statistical investigations as part of the routine. Many researchers have reported convincingly that in the presence of certain kinds of tasks using technological tools, younger and less statistically trained learners appear quite adept at reasoning in quite sophisticated ways (e.g., Ainley, 2000; Fitzallen, 2012; Lehrer, Kim, & Schauble, 2007; Manor & Ben-Zvi, 2015; Watson & Donne, 2009). These studies add strength to Pea's (1985) suggestion that technology may provide cognitive tools that allow learners to interact with seemingly complex ideas in ways that would be impossible without the tools.

Dynamic technological tools such as *TinkerPlots* (Konold & Miller, 2015) and *Fathom Dynamic Data Software* (Finzer, 2012) have received significant attention as tools of choice for many researchers working with students and teachers. These tools allow users to generate, import, or input data; to conduct simulations, model, and test hypotheses; and to construct important linkages across representations. Unlike many simulation environments where students may simply engage at a distance, an environment like *TinkerPlots* allows users to explore representations quickly. Perhaps most importantly, and one reason for its great appeal, *TinkerPlots* was designed to incorporate students' intuitive notions of data and representations and to minimize the distance, from what students' initial thoughts and representational preferences might be, to those available in the tool (Konold, 2007). For an extended discussion of *TinkerPlots* and *Fathom* features, see Biehler, Ben-Zvi, Bakker, and Makar (2013) and Watson and Fitzallen (2016).

At the time of this publication, many additional visualization and simulation tools are currently available for supporting representational work (see Table 4.1), but the extent to which these new tools are utilized in classrooms is unknown. These tools range on the continuum from exploratory modeling tools where users explore expert built models, to expressive modeling tools, where students can construct their own models (Doerr & Pratt, 2008). As such, they provide different opportunities for learners to control the technology and therefore express their representational preferences and flexibility by using different representations.

The usefulness of a tool to support learning likely depends on the extent to which the learner is actively engaged in the construction of objects and relationships while representing and modeling. It is quite plausible for a learner to manipulate a representation while not attending to the more generalizable features of the system

Table 4.1 Online data analysis and visualization tools

Tool	URL	Description
Building Concepts: Statistics and Probability	http://www.tibuildingconcepts.com/activities/statistics	Learning sequences incorporating the TI-NSpire calculator to support the development of statistical concepts
CODAP	http://codap.portal.concord.org	Web-based data tool designed as a platform for developers and as an application for students in Grades 6–14
Core-Math Tools	http://www.nctm.org/coremathtools/	Downloadable suite of interactive software tools for algebra and functions, geometry and trigonometry, and statistics and probability; Java required
Desmos	https://www.desmos.com	Digital math tools including HTML5 Desmos graphing calculator and activities built on top of that calculator; includes an Activity Builder, helping teachers create digital math activities
Fathom	http://fathom.concord.org	Dynamic software that is fun and effective for teaching data analysis and statistics; also a powerful tool for high school students to use for modeling with mathematics
GapMinder	http://www.gapminder.org/world/	Independent Swedish foundation that produces free teaching resources making the world understandable based on reliable statistics
GeoGebra	https://www.geogebra.org/home	Mathematics calculators with graphing, geometry, 3D, spreadsheet, CAS, and more
Interactivate	http://www.shodor.org/interactivate/activities/	Free, online courseware for exploration in science and mathematics comprised of activities, lessons, and discussions
iNZight	https://www.stat.auckland.ac.nz/~wild/iNZight/index.php	Tools for students to quickly and easily explore data and understand statistical ideas; extends to multivariable graphics, time series, and generalized linear modeling
National Library of Virtual Manipulatives	http://nlvm.usu.edu	Digital library containing Java applets and activities for K-12 statistics and mathematics
Plotly	https://plot.ly/feed/#sob	Collaboration platform for modern data science

(continued)

Table 4.1 (continued)

Tool	URL	Description
Quandl	https://www.quandl.com	Designed for professionals, delivers financial, economic, and alternative data to over 200,000 people worldwide
Rice Virtual Lab in Statistics	http://onlinestatbook.com/stat_sim/index.html	Repository including HyperStat Online, simulations, demonstrations, case studies, and analysis lab
Rossman/Chance Applet Collection	http://www.rossmanchance.com/applets/	Set of interactive simulations for exploring statistical concepts from data analysis to bootstrapping
Stat Key	www.lock5stat.com/StatKey	Powerful JavaScript simulations to support *Statistics: Unlocking the Power of Data* text
TinkerPlots	https://www.tinkerplots.com	Dynamic data exploration, visualization and modeling tool developed for use by middle school through university students
Tuva Labs	https://tuvalabs.com	Authentic data sets and tools to provide students opportunity to apply mathematics and statistics concepts to solve meaningful problems in real-world contexts
RStudio	https://www.rstudio.com	Open source, integrated development environment for R; includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging, and workspace management

(Fitzallen, 2013). This could explain the disappointment that some researchers have faced when discovering a lack of learning demonstrated by their students when using simulations in learning environments (Chance, delMas, & Garfield, 2004; Lane & Peres, 2006; Mills, 2002).

Students are often introduced to graphical representations using physical manipulatives and embodiments, such as Post-it Notes, objects, or counters (Chick & Watson, 2001; Friel, Curcio, & Bright, 2001). In some cases, students create representations as a matter of providing a record of the data, while in other cases, they create representations "with the hope that information that is not otherwise apparent will emerge from them" (Monk, 2003, p. 252). Early research identified challenges associated with novice graphical creation, interpretation, and competency (cf. Friel et al., 2001). "However, when the constraint of drawing graphs by hand is removed, primary age children are able to utilize computers to work with line and scatter graphs long before they would typically meet them in the school curriculum" (Ainley, 2000, p. 368). Even with technological tools, inviting novice learners to generate meaningful representations with physical materials is still recommended as an entry point; however, the debate about when and how to utilize technology to augment or offload the representation burden for learners is ongoing and under-researched. In school settings, there has been a long history of teaching statistics focused on the construction of privileged representations, now including line plots, dot plots, pie graphs, stem-and-leaf displays, box plots, bar graphs, histograms, and scatterplots. What is required is more flexibility in the curriculum, followed by research to explore various contexts and their relationship to creative representations about them (Monk).

Prior to instruction, students often create what are referred to as case value plots, collections of bars of lengths corresponding to the magnitude of individual cases (Bakker, 2004; Cobb & McClain, 2004). *TinkerPlots* offers a case value plot option that allows students to build understanding from this familiar representation and extend it easily to other forms, thus facilitating connections (see Fig. 4.3d). Hat plots, also in *TinkerPlots*, are newer representations that resemble box plots without the median line and where the whiskers are level with the bottom of the box, as shown in Fig. 4.2. Researchers have found hat plots useful to highlight the center clump in a distribution while avoiding some of the pitfalls experienced with box plots (Bakker, Biehler, & Konold, 2005; Watson, Fitzallen, Wilson, & Creed, 2008).

Consider the potential technological, pedagogical, and statistical demands associated with the construction and interpretation of images in Fig. 4.3. Each figure represents data from an experiment modeling 50 flips of a six-sided die to estimate the mean roll of the die. All of the representations support *seeing the data* in ways students are expected to experience, but each provides a view that illuminates some things and hides others. Within and across representations, deep connections can be made to support strong conceptual understanding of the mean as a useful measure, providing a fruitful area for research.

When thinking about the use of technology, it is also important to consider who is using the technology and how they are using it (Fitzallen, 2013; Madden, 2013; Trouche, 2005). With the evolution of new tools, it is important to continue to conduct research on the nature of learners' engagement and learning with the tools

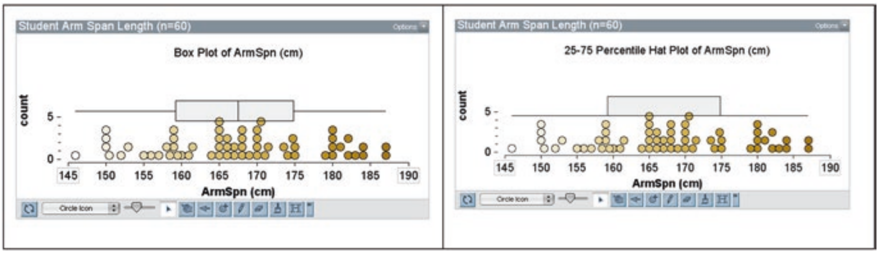


Fig. 4.2 Subtle difference between box plot (*left*) and hat plot (*right*)

as they conduct statistical investigations. Longitudinal studies have demonstrated that children as young as early elementary school can be quite capable when using visual representations to reason about variation and expectation (e.g., English, 2010, 2012), but it is fair to say that most students do not yet benefit from these types of opportunities to learn (Lehrer et al., 2007). Further discussion of reasoning about distributions is found in Chap. 5.

4.3.4 Data Analysis: Summarizing and Reducing Data

Continuing from visual representations, sometimes there is the need to summarize or reduce data in order that the story in the data is amplified. Amplification can be achieved by reducing the data to measures of center and spread or through the use of graphical summary representations, such as the box plot. This process supports students in answering questions by allowing them to see the aggregate characteristics of data, such as center, spread, and shape, which are not evident in any of the individual cases (Konold & Higgins, 2003).

Much of the early research in statistics education was focused on central tendency and how students conceptualized the concept in terms of representativeness, location, and expectation (e.g., Goodchild, 1988; Strauss & Bichler, 1988). It was motivated by the earlier work of Pollatsek, Lima, and Well (1981) who found college students had difficulties using effective computational skills to determine weighted means. Goodchild found that the students could calculate the mean but had underdeveloped notions of expectation and representativeness. Strauss and Bichler placed less emphasis on mathematical calculations and focused on students' understanding of properties of the mean and their development over time.

In 1990, Russell and Mokros conducted a study with 21 each of Grade 4, 6, and 8 students, who responded to seven construction and interpretation problems. Similar to the results reported by Strauss and Bichler (1988), Russell and Mokros reported that the students in their study were able to calculate the mean and used a variety of strategies to solve central tendency problems. They reported that students conceptualized the "average" as: "(1) average as modal; (2) average as what's average;

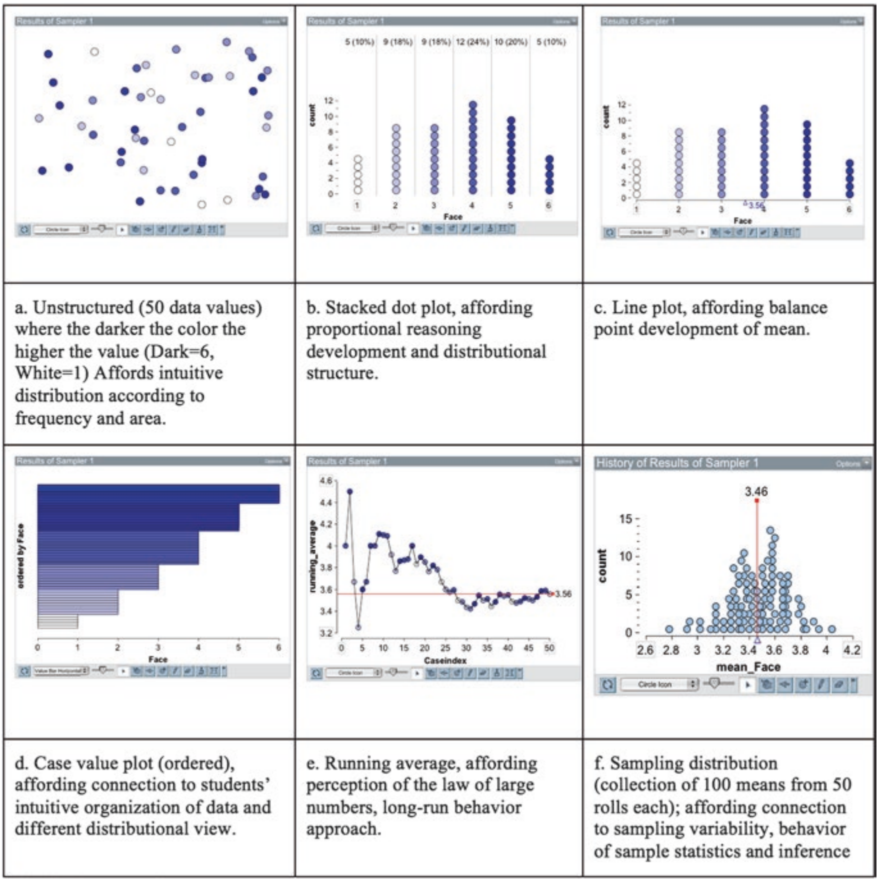


Fig. 4.3 Representations related to modeling 50 rolls of a six-sided die. Darkest color represents 6; white represents 1. Graphs (**a**)–(**e**) are different representations of the same 50 values; graph f is a collection of 100 means from simulating 50 rolls of a die 100 times. 3.46 is the mean of the empirical sampling distribution

(3) average as the midpoint; and (4) average as an algorithmic relationship" (p. 308). In 1995, they extended the list to include average as a point of mathematical balance (Mokros & Russell, 1995). Similar ideas about central tendency emerged from a study of Grade 3 students by Makar (2014). These students, who had not been introduced to the mean algorithm, described central tendency in the context of measuring the heights of students, as (1) a reasonable height, (2) the most common value, (3) the middle height, (4) the normal height, and (5) representative of the population.

Exploring students' conceptual understanding of the mean algorithm was the focus of a large study of 250 Grade 6 students conducted by Cai (1998). Most students recognized the mean algorithm, but only half were able to apply the concept to solve open-ended questions. Watson and Moritz (1999) reported similar results from their longitudinal study of Grade 3–11 students. Older students could apply the mean algorithm but did not use its representative property to compare data sets. Watson and Moritz offered a model of development of concepts of central tendency based on the structure of observed learning outcomes (SOLO) model (Biggs & Collis, 1982), with levels based on the complexity of the structure of responses—prestructural, unistructural, multistructural, and relational. Watson and Moritz (2000b) extended the model to higher levels based on interviews with 94 students, when more complex questions were asked.

Researchers have often recommended that graphical representations be used to foster and build students' understanding of central tendency (e.g., Konold & Harradine, 2014; Leavy, Friel, & Mamer, 2009; Lehrer, Kim, & Jones, 2011). Bakker, Derry, and Konold (2006), however, note the difficulties students have understanding the characteristics of graphical representations for this purpose, in particular box plots. Although box plots reduce data to a five-point summary, Bakker et al. suggest that students have difficulty understanding the meaning of the representation because the size of the sections of a box plot is inversely related to the density of the data. Seeing central tendency in terms of an interval, however, is worthwhile, and further research with the hat plot may show it to be a useful tool (e.g., Watson et al., 2008).

When analyzing graphical representations, students may intuitively look at the shape of the distribution and be drawn to clusters and clumps of data to make decisions (Bakker et al., 2006; Konold, Higgins, Russell, & Khalil, 2015). During a study that looked at the distribution of the size of genetically modified fish and normal fish in a population, Bakker et al. reported that Grade 6 students used the notion of "clump" to refer to the majority of values in the middle of the graph before moving on to use the mean as a group descriptor to identify the middle clump of the distribution. They then used the formal mean value to compare the distributions of the two types of fish. Konold and Harradine (2014) also reported students aged 12–14 identifying modal clumps as indicators of center before they used formal measures such as the mean when investigating a manufacturing process.

Following the work of Bakker et al. (2006), recent studies have also been exploring students' understanding from broader perspectives. Complementing this are suggestions from Konold and Harradine (2014), who claim that repeated measurements and production processes are particularly fruitful contexts for introducing students to statistical ideas about variation and measures of center. Makar (2014) also illustrated how basing the development of informal inference within inquiry-based learning experiences can support young students to develop rich conceptions of central tendency. As well in 2014, Watson, Chick, and Callingham considered the context within which tasks on central tendency were presented to 247 middle school students in surveys. They found differences in performance across contexts and a peak in performance at Grade 9. These studies illustrate the possibilities for research linking central tendency more closely to other stages of the Practice of Statistics.

4.3.5 *Drawing Conclusions: Decision-Making and Informal Inference*

Historically, the development of statistical knowledge and decision-making in the senior school years has been attempted predominantly through the presentation of formal statistics, with little or no recognition of the concept of informal inference. Meanwhile, students in the primary and early secondary years have typically learned about descriptive statistics, such as measurements of center, and how to produce and interpret a limited range of graphs at a fairly simple level. Furthermore, the content has usually been taught in a piecemeal manner, where statistics and probability have been treated as separate topics (Bakker & Derry, 2011). Such a focus on skills, procedures, and computations does not lead students to develop statistical reasoning and thinking (Makar & Ben-Zvi, 2011) because they often do not use the tools to address statistical problems (Bakker & Derry). At a higher level, pre-tertiary and tertiary statistics courses usually begin to address inferential statistics through the derivation, application, and/or interpretation of theoretical models. Such courses may also develop students' mechanistic application of methods (Ben-Zvi & Garfield, 2004) as distinct from knowledgeable decision-making, thereby doing little to advance the statistical reasoning and literacy of those students who do engage in higher-level courses (Zieffler et al., 2008).

The difficulties of studying formal statistical inference at high school and tertiary levels have been documented in research (e.g., Garfield & Ben-Zvi, 2008) with the suggestion that problems may stem from students' inexperience with the underpinnings of statistical inference (Pfannkuch, 2005). Pratt, Johnston-Wilder, Ainley, and Mason (2008) suggest that informal approaches to inference could, and should, underpin pre-formal statistics education. To have students take such an approach requires an appreciable shift in the way that students address statistics. Fortunately, there has been a surge in research in the last decade to address how students can be supported to engage with inference informally.

As noted earlier in relation to frameworks, the research on informal inference is now suggested as the process underpinning the entire Practice of Statistics at the school level before formal theories are employed (cf. Chap. 8). Using the phrase "informal inference" in relation to other frameworks emphasizes a distinction in the type of evidence collected and the type of analysis performed. The focus on uncertainty related to the sample-population relationship is implicit in the other frameworks, but research may show that making it more obvious results in greater understanding by students. It may also foster greater transfer of the fundamental requirement for uncertainty in later encounters with statistics.

An informal inferential approach serves to make the development of statistical underpinnings available from very early stages, for example, with emergent inferential practices being developed with 5-year-olds (Makar, 2016) and with 9-year-olds (Watson & English, 2015). One of the most significant contributions enabling advancements in informal reasoning has been progress made in technology that enables data visualization and manipulation at an early age, for example, with *TinkerPlots* (Watson & Fitzallen, 2016), as well as randomization procedures for older students, for example, with *iNZight* (Budgett, Pfannkuch, Regan, & Wild, 2013).

4.3.6 The Importance of Learning Through Complete Investigations

Having considered the components of the Practice of Statistics, it is relevant to consider research on combining the components in a complete investigation. Most consultant statisticians are provided with a context and a question or questions to answer when employed to carry out their Practice of Statistics. Is it then realistic to ask school students to find meaningful contexts and questions that lead to the application of the statistical tools with which they are familiar? Although Chin and Kayalvizhi (2002) and Lavigne and Lajoie (2007) gave considerable freedom to students to pose questions, the inclusion of scaffolding examples influenced the results. Accepting the dictate of Rao (1975) that there can be no statistics without context, it seems reasonable to begin statistical investigations at school with a meaningful overall topic for inquiry. This is also practical for the classroom teacher in terms of personal expertise and time constraints.

Accepting that the Practice of Statistics begins within an agreed context, the question then becomes related to the age at which exposure to the Practice of Statistics should occur. The National Council of Teachers of Mathematics (NCTM) in the United States in its 1989 *Standards* for early childhood set a firm foundation.

For grades K-4, the mathematics curriculum should include experiences with data analysis … so that students can:

- Collect, organize, and describe data.
- Construct, read and interpret displays of data.
- Formulate and solve problems that involve collecting and analyzing data (p. 54).

Twenty years later, the US *Common Core State Standards: Mathematics* (CCSSI, 2010) ignored consideration of statistical topics until Grade 6 and introduced statistical investigations in Grade 7. At this time, however, the New Zealand *Mathematics and Statistics* curriculum (Ministry of Education, 2009) was renamed, and "Statistical investigation" was one of the three subheadings for Statistics at every level of the curriculum. At Level 1, students:

Conduct investigations using the statistical enquiry cycle:

- Posing and answering questions
- Gathering, sorting and counting, and displaying categorical data
- Discussing the results

The research of Fielding-Wells (2010), however, suggests that students experience considerable difficulty in envisioning and understanding the power and potential of statistics when addressing the initial stages of an investigation. This result may be an indication that learning experiences should initially expose students to the later stages of the Practice of Statistics where they can experience the purpose and success of reaching a decision on a question posed by others. This could be the basis for future research programs.

In a classroom environment, it is likely that, even when setting a well-defined context, the entire process from posing a question through drawing a conclusion, completing the entire investigation, will be very time-consuming on its own. One of the few studies claimed to report on an entire statistical investigation is that of Lavigne and Lajoie (2007). The authors also followed the students' data collections, data analyses, and data representations, identifying ten types of reasoning potentially present in each of the four phases of the inquiry. In a study with Grade 5 students, Watson and English (2015) set a general question about deciding if different populations were environmentally friendly. Students contributed to the problem posing by setting criteria for being environmentally friendly based on five questions about behavior, such as having shorter showers and recycling rubbish. They collected and analyzed data first from their class and then from random samples taken from a national database of Grade 5 students. Based on the evidence from their samples, students drew informal inferences for the different populations of Grade 5 students, justifying their confidence in the decision each time.

A somewhat different hierarchical classification for analyzing students' work as they progress through an investigation is suggested by Fitzallen, Watson, and English (2015). Growing out of the SOLO model, the levels are assigned for each stage of the Practice of Statistics. The investigation as a whole is then classified by how the outcomes of the four stages are combined to make a decision. Another recent approach is that of Fielding-Wells and Makar (2015) who used an argumentation framework stressing the need for evidence to build understanding and to analyze students' development of informal inference acknowledging uncertainty. This area is one that requires continuing research, which although complex and time-consuming, should be very rewarding. One approach is to ask students to write a final report, describing every stage of the investigation (Forster & Wild, 2010). This focus on report writing needs to receive attention at lower levels as well where students often give oral presentations on their findings (English, 2015).

The realization that the planning and facilitating of complete investigations at school level can be onerous has led to research on helping teachers implement such programs. In this situation, teachers may be only marginally ahead of their students in terms of understanding the content and the implementation of the Practice of Statistics (Burgess, 2011). Based on case studies with teachers, Burgess considers types of content and pedagogical knowledge needed by teachers in order to use statistical investigations in the classroom in relation to the PPDAC cycle. Santos and Ponte (2014) present a detailed case study of a preservice primary teacher's experience in learning about statistical investigations and implementing one in a

Grade 3 class. Working through the PPDAC cycle, the authors illustrate the value of exploring teachers' understanding in depth and the many issues for beginning teachers with limited experience themselves. The rewards and difficulties of 23 in-service teachers during their first inquiry-based lessons with their students following professional development are explored by Makar (2010). She concludes that the initial experience can be frustrating due to the uncertainty of data outcomes, the logistics of the classroom activity, and content concerns. Makar and Fielding-Wells (2011) make specific suggestions, based on other research and their own, for assisting teachers at each of the five stages of the PPDAC cycle. At the secondary level, Madden (2011) suggests introducing teachers to provocative tasks to motivate complete investigations and informal inference. Batanero, Burrill, and Reading (2011) present insights that could be the basis for further research related to the teacher educator's role in assisting teachers to handle the Practice of Statistics in the classroom.

Finally, in thinking about carrying out an investigation, it is important to acknowledge the potential for occasional or even frequent backtracking depending on the obstacles or revelations met along the way. Often, the final report on a statistical investigation does not report details about the backtracking, but only about the successful path followed through the investigation, which may give students a false sense of security about the ease with which they can carry out the Practice of Statistics. Students hence are likely to benefit from experiencing complete statistical investigations themselves: to feel the uncertainty, the occasional frustration, the necessity to rethink at various points, perhaps the necessity to backtrack to rethink the questions or represent the data differently, and finally the task of writing a reasonable report on the entire investigation and its outcome. Although Konold and Higgins (2003) address issues of looking forward and backward while carrying out an investigation, no research was found on focusing specifically on students experiencing this phenomenon.

As there are very few classroom studies based on students carrying out a complete statistical investigation, there are many opportunities and questions for researchers to consider about the Practice of Statistics at school. Are some stages of the practice more difficult than others? If so, which are they and why are they more difficult? Does difficulty depend on the context, implying that the most difficult stages are different for different contexts? Is long-term retention greater for the stages of the practice if they have been embedded in a meaningful context than if they have been taught in an isolated manner? Does the exposure to complete investigations and drawing informal inferences at school build the appropriate foundation for formal inference when introduced at the college level? Given the difficulties some tertiary students have in interpreting *p*-values (e.g., Reaburn, 2014), it might be expected that the earlier experience would provide a meaningful conceptual framework when the more sophisticated tools are employed.

4.4 The Big Ideas

As seen in this chapter, there is general agreement, with minor variations, on the purpose of the Practice of Statistics being to solve problems by conducting investigations. In this potentially complex environment, the foundational concepts that underpin the Practice of Statistics in carrying out a statistical investigation are also a focus of researchers' attention. Often described as Big Ideas, there is some debate on how many of these fundamentals there are.

In the summary to their book on statistical literacy, reasoning, and thinking, Ben-Zvi and Garfield (2004) presented a list of eight Big Ideas adapted from the work of Susan Friel. The eight Ideas were data, distribution, trend, variability, models, association, samples and sampling, and inference (p. 400). In a later work, Garfield and Ben-Zvi (2008) added a ninth Big Idea, Comparing Groups. Burrill and Biehler (2011) did not use the adjective "big" but introduced seven "fundamental" statistical ideas as critical for teachers to know and convey. These agree well with the Big Ideas of Ben-Zvi and Garfield. Big Ideas were also a feature of the NCTM books on essential understanding of statistics for Grades 9–12 (Crites & St. Laurent, 2015; Peck, Gould, & Miller, 2013). Rather than single terms or phrases, the five Big Ideas were couched in sentences summarizing a total of 24 essential understandings:

Big Idea 1. Data consist of structure and variability.
Big Idea 2. Distributions describe variability.
Big Idea 3. Hypothesis tests answer the question, "Do I think that this could have happened by chance?"
Big Idea 4. The way in which data are collected matters.
Big Idea 5. Evaluating an estimator involves considering bias, precision, and the sampling method (Crites & St. Laurent, 2015, pp. 127–128).

In 2013, Watson, Fitzallen, and Carter were asked by the Australian Association of Mathematics Teachers to present five (only) Big Ideas for teaching and learning statistics in Grades 6–10. This restriction resulted in the five ideas shown in Fig. 4.4, which were an attempt to provide foundations for planning and carrying out a statistical investigation before the introduction of formal inference. In doing so, the figure shows all of these Big Ideas as closely related throughout the Practice of Statistics; it is not based on stages in which an investigation is carried out. In fact, variation is an influence at every stage of the Practice of Statistics.

The claim that variation is the "most fundamental" of the Big Ideas (Moore, 1990) is now recognized in the school curricula of Australia in Foundation to Year 2 (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2015), of New Zealand at Level 4 (Ministry of Education, 2009) and of the United States in Grade 6 (CCSSI, 2010). At the middle school level, the term expectation, as used in Fig. 4.4, reflects both the potential outcomes of data analyses, perhaps initially expressed as hypotheses, and the elementary probabilities of chance outcomes. Historically, in the curriculum and classroom, expectation was encountered first in measures of center before variation in the guise of standard deviation (Shaughnessy, 1997). Watson's research (2005, 2009), however, suggests that children develop an appreciation of variation before an appreciation of expectation (Fig. 4.4). Shaughnessy (2006) reflects this foundational status of variation and expectation in his choices of the *two* Big Ideas in statistics arising from the work of Wild and Pfannkuch (1999). Wild (2006) provides the essential background on the importance of distribution (cf. Fig. 4.4) as the lens through which variation in data is viewed and hence analyzed. Distributions provide ways of visualizing data that allow for decisions to be made about questions related to expectation. Randomness arises in processes that have unpredictable individual outcomes but display patterns over the long term (cf. Moore, 1990). Randomness in samples of sufficient size and incorporated in experimental design is the basis for increased confidence in informal inferences. At the school level, informal inference encompasses the type of decision-making that occurs in statistics (cf. Chap. 8).

With this fundamental acknowledgement of Big Ideas, the Practice of Statistics becomes more than a series of procedures to reach a decision. One of the future directions of research is to devise studies that not only involve diverse procedures for the Practice of Statistics but also help students understand and appreciate the Big Ideas behind them.

4.5 Statistical Literacy: Assessing the Claims of Others

Until this point, Chap. 4 has been about the actual Practice of Statistics, *doing* statistics. In situations outside of the school classroom, involvement with statistics is not always, or perhaps even often, about practicing statistics but about judging the

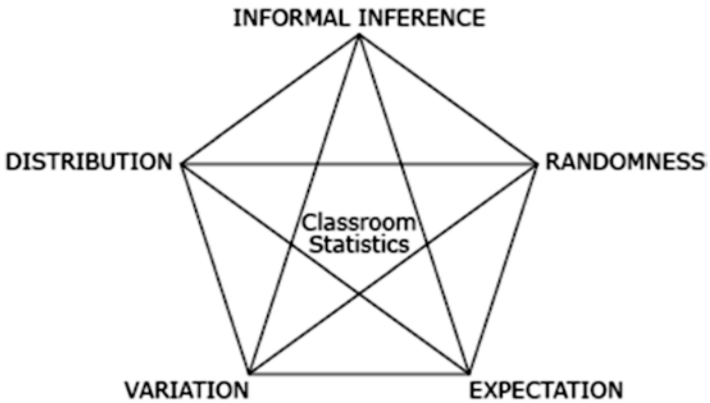


Fig. 4.4 Interrelated Big Ideas underlying statistics in the classroom (Watson, Fitzallen, & Carter, 2013)

outcomes and claims of others who practice statistics. In most situations, one has to decide if data are sound, if graphs presented are appropriate and correct, if the "average" chosen suits the context, and/or if the decision reached is believable. One also sometimes needs to explain what is wrong and why. This process requires the ability to argue effectively, an aspect that has received recent attention in relation to informal decision-making. As formulated by Toulmin, Rieke, and Janik (1984), argumentation takes two forms: inquiry and advocacy. Fielding-Wells and Makar (2012, 2015) and Makar, Bakker, and Ben-Zvi (2015) focus on argumentation as part of inquiry involving informal inference. Assessing claims of others, however, is likely to require argumentation associated with advocacy. This involves recognizing the advocacy in the claim and the ability to advocate with evidence for an alternate claim.

As Gal (2002) states for adults, statistical literacy is:

(a) The ability to *interpret and critically evaluate* statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant
(b) The ability to *discuss or communicate* their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions (p. 2)

The assumption may be that after one has completed an education based on the content and pedagogy described throughout this chapter, one would be statistically literate in Gal's sense. Experiencing the "correct" way to practice statistics, however, may not train one to be a detective in other contexts, to know what questions to ask, and to be able to explain the difference between what is observed and what is legitimate.

The close relationship of statistical literacy to the Practice of Statistics itself, however, has meant that on some occasions, the distinction becomes blurred. Although perhaps taught side by side, there must be awareness by curriculum designers and teachers of the requirements of each. This is recognized in the school curriculum of New Zealand where *Mathematics and Statistics* (Ministry of Education, 2007) has within the Statistics section at every level a subheading of "statistical literacy" at the same standing as the subheading for "statistical investigation." Although not as extensive as statistical investigation, statistical literacy acknowledges the importance of students being able to apply their understanding in various contexts. In Australia, the *General Capabilities* (ACARA, 2013) expected of all students across the curriculum, under numeracy, include an element, "interpreting statistical information." Although not using the word "literacy," the description fits closely that of statistical literacy. In the United States, Statistical literacy is not specifically mentioned in the Standards for Mathematical Content section of the *Common Core* (CCSSI, 2010). In the Standards for Mathematical Practice section, however, the third listed practice is "construct viable arguments and critique the reasoning of others." Within the description, students:

> reason inductively about data, making plausible arguments that take into account the context from which the data arose … [They] are also able to compare the effectiveness of two plausible arguments, distinguish correct logic or reasoning from that which is flawed, and if there is a flaw an argument—explain what it is. (pp. 6–7).

A framework for learning, applying, and assessing statistical literacy is suggested by Watson (2006). The three tiers of statistical literacy form a hierarchy:

Tier 1: Understanding the terminology to be used for statistical decision-making
Tier 2: Interpreting the terminology in the context presented
Tier 3: Possessing the ability and confidence to challenge statements made in the context without proper statistical foundation

Claims requiring Tier 3 assessment may be correct or incorrect, which is why developing critical thinking skills is important. The use of the hierarchy for assessing statistical literacy is illustrated by Watson and Moritz (2000a) for survey items related to sampling. This framework, particularly Tier 3, provides settings for research on argumentation related to advocacy, as noted earlier. This could be a fruitful area of future research given the frequent appearance of motivating contexts in the media.

Watson and Callingham (2003) used 80 items and a sample of over 3000 students to suggest a general six-level hierarchy of understanding with two levels approximating each of the three tiers of the statistical literacy hierarchy. Some of the items used can also assess understanding associated with the data collection stage of carrying out a statistical investigation, but they do not cover all aspects of the Practice of Statistics. Watson and Callingham included items in their analysis that assessed probabilistic literacy as well as statistical literacy as in most school curricula they are considered under the same heading (e.g., ACARA, 2015; CCSSI, 2010; Ministry of Education, 2009). Gal (2005), however, has suggested some aspects of probabilistic literacy that distinguish it from statistical literacy, particularly noting independence. The recent work in the LOCUS (Levels of Conceptual Understanding in Statistics) project has the aim of creating items to assess all aspects of the Practice of Statistics from a statistical literacy perspective (Whitaker, Foti, & Jacobbe, 2015). As implied in the project's name, the assessment items measure conceptual, rather than procedural, understanding based on the *GAISE* framework. Using both open-ended, as well as multiple-choice questions, assessments are more comprehensive than the surveys of Watson and Callingham (2003) in their reflection of the complete Practice of Statistics.

Beyond the classroom, the contexts that require statistical literacy vary tremendously. This has led to many different approaches to research in the area. In 2002, the Sixth International Conference on the Teaching of Statistics, ICOTS6, included a keynote address, a panel discussion, and six papers in a special session on statistical literacy. Opinions differed on the diversity in and kinds of statistical literacy (Murray & Gal, 2002; Schield, 2002), on what citizens should know (Moreno, 2002; Utts, 2002), and on how to reach the level required (Boland, 2002; Botting & Stone, 2002; Phillips, 2002). It was 12 years later when ICOTS9 for the first time had an entire topic devoted to statistical literacy in the wider society, with five subtopics, including requirements (e.g., Hovermill, Beaudrie, & Boschmans, 2014) and assessment (e.g., Bidgood, 2014), as well as studies following student development (e.g., Sproesser, Kuntze, & Engel, 2014). Outside of the specific topic, there were papers featuring statistical literacy within other topics focusing on the workplace specifically, reporting on international collaborations to assist in-service teachers (e.g., North, Zewotir, & Gal, 2014), relating progress in different countries (e.g., MacFeely & MacCuirc, 2014), and specializing in areas such as risk literacy (e.g., Till, 2014) or applying knowledge reading scientific journals (e.g., Esfandiari, Sorenson, Zes, & Nichols, 2014). The variety of research reports from ICOTS not only shows topics currently under scrutiny but also opens the imagination to many future research opportunities.

In this section, it has been suggested that students' engagement in the Practice of Statistics at school might be sufficient to produce statistically literate citizens. There is also the question, however, as to whether it is necessary. Is it possible to become statistically literate without hands-on experience, perhaps by reading good books or viewing instructional videos on the topic? This is an unanswered research question for which it would take well-designed research spanning many years to answer.

4.6 Summary

The themes of this chapter provide an avenue into Section II of the Handbook from the perspective of the Practice of Statistics. After summarizing several useful frameworks for carrying out the practice, research on five stages was introduced in more detail. The areas with the greatest possibilities for further research appear to be posing and refining questions, planning and collecting data, and drawing conclusions while acknowledging uncertainty. Research could fruitfully consider the importance of the stage at which students are initially exposed to the Practice of Statistics. Should students start at the *very* beginning by posing questions or first experience later stages in order to appreciate the purpose of the investigative process?

Does introducing the Practice of Statistics at the school level with motivating contexts and stimulating questions have the potential to improve the attitudes of students to a subject that has traditionally received "bad press" at more senior levels? Experiencing informal inference with the type of evidence from data available at the school level and appreciating uncertainty in decision-making are important contributors both to being statistically literate citizens and to understanding formal statistical inference at the tertiary level. The confidence gained from embracing the Practice of Statistics at school can be hypothesized to provide a concrete foundation for theoretical tertiary study. Research over a long period of time would be needed to support the claim!

Perhaps the most contentious question, raised at the end of the previous section, facing researchers and curriculum designers is, "Can students leave school statistically literate without having experienced the Practice of Statistics during their years of schooling?" Is the experience necessary? The authors of the New Zealand curriculum (Ministry of Education, 2007) may have solved the problem by presenting the topics "statistical investigation" and "statistical literacy" side by side across the years of schooling. If all students experience the Practice of Statistics as well as being exposed to statistical literacy, then the research question would appear to become one of sufficiency. Given the concerns of Zieffler et al. (2008) at the tertiary level, however, a call for research across the years of learning seems to be appropriate.

The Big Ideas underlying the Practice of Statistics continue to evolve as research expands across the levels of education where they are encountered. It is essential that research related to the Practice of Statistics, in whatever form it is introduced, include understanding of the fundamental concepts as well as carrying out the procedures. Many more specific research questions arise about the Practice of Statistics in the rest of the chapters in this Handbook. Although there has been a surge of research in the last 20 years, there are still many more puzzles to solve related to statistical thinking.

References

Ainley, J. (2000). Transparency in graphs and graphing tasks: An iterative design process. *The Journal of Mathematical Behavior, 19*(3), 365–384.

Allmond, S., & Makar, K. (2010). Developing primary students' ability to pose questions in statistical investigations. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society (Proceedings of the 8th International Conference on the Teaching of Statistics, Ljubljana, Slovenia, July 11–16)*. Voorburg, The Netherlands: International Statistical Institute.

Arnold, P. (2008). What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation. *Proceedings of the 11th International Congress of Mathematics Education, Monterrey, Mexico, 6–13 July 2008*.

Australian Bureau of Statistics. (2011). *CensusAtSchool Australia: Your random data sample.* Retrieved from www.cas.abs.gov.au/cgi-local/cassampler.pl

Australian Curriculum, Assessment and Reporting Authority. (2013). *General capabilities in the Australian Curriculum, January, 2013*. Sydney: Author.

Australian Curriculum, Assessment and Reporting Authority. (2015). *The Australian curriculum: Mathematics, Version 8.0, August 21, 2015*. Sydney: Author.

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64–83.

Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about box plots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Education (IASE) Roundtable* (pp. 163–173). Voorburg, The Netherlands: International Statistics Institute.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1 & 2), 5–26.

Bakker, A., Derry, J., & Konold, C. (2006, July). Using technology to support diagrammatic reasoning about center and variation. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education (Proceedings of the 7th International Conference on the Teaching* *of Statistics, Salvador, Bahia, Brazil, July 2–7)*. Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute.

Batanero, C., Burrill, G., & Reading, C. (Eds.). (2011). *Teaching statistics in school mathematics—Challenges for teaching and teacher education: A joint ICMI/IASE study*. Dordrecht, The Netherlands: Springer.

Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics, 88*(3), 291–303.

Ben-Zvi, D., & Garfield, J. (Eds.). (2004). *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, The Netherlands: Kluwer.

Bidgood, P. (2014). Towards statistical literacy—Relating assessment to the real world. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13–18)*. Voorburg, The Netherlands: International Statistical Institute.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–690). New York: Springer.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Boland, P. J. (2002). Promoting statistics thinking amongst secondary school students in the national context. In B. Phillips (Ed.), *Developing a statistically literate society (Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town, South Africa, July 7–12)*. Voorburg, The Netherlands: International Statistical Institute.

Botting, B., & Stone, D. (2002). Experience of dealing with the media on congenital anomaly research. In B. Phillips (Ed.), *Developing a statistically literate society (Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town, South Africa, July 7–12)*. Voorburg, The Netherlands: International Statistical Institute.

Bright, G. W., & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 63–88). Mahwah, NJ: Lawrence Erlbaum.

Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. J. (2013). Dynamic visualizations and the randomization test. *Technology Innovations in Statistics Education, 7*(2), Article 5.

Burgess, T. A. (2011). Teacher knowledge of and for statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 259–270). Dordrecht, The Netherlands: Springer.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 57–69). Dordrecht, The Netherlands: Springer.

Cai, J. (1998). Exploring students' conceptual understanding of the averaging algorithm. *School Science and Mathematics, 98*(2), 93–98.

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295–323). Dordrecht, The Netherlands: Kluwer.

Chick, H. L., Pfannkuch, M., & Watson, J. M. (2005). Transnumerative thinking: Finding and telling stories with data. *Curriculum Matters, 1*, 87–108.

Chick, H. L., & Watson, J. M. (2001). Data representation and interpretation by primary school students working in groups. *Mathematics Education Research Journal, 13*(2), 91–111.

Chin, S., & Kayalvizhi, G. (2002). Posing questions for open investigations: What questions do pupils ask? *Research in Science & Technology Education, 20*(2), 269–287.

Cobb, P., & McClain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 375–395). Dordrecht, The Netherlands: Kluwer.

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association for Best Practices and the Council of Chief State School Officers.

Crites, T., & St. Laurent, R. (2015). *Putting essential understanding of statistics into practice in grades 9–12*. Reston, VA: National Council of Teachers of Mathematics.

Doerr, H. M., & Pratt, D. (2008). The learning of mathematics and mathematical modeling. In M. K. Heid & G. W. Blume (Eds.), *Research on technology and the teaching and learning of mathematics: Volume 1 Research syntheses* (pp. 259–286). Charlotte, NC: Information Age Publishing.

English, L. D. (2010). Young children's early modelling with data. *Mathematics Education Research Journal, 22*(2), 24–47.

English, L. (2012). Data modeling with first-grade students. *Educational Studies in Mathematics, 81*(1), 15–30.

English, L. D. (2015). STEM: Challenges and opportunities for mathematics education. In K. Beswick, T. Muir, & J. Wells (Eds.), *Climbing mountains, building bridges* (*Proceedings of 39th conference of the International Group for the Psychology of Mathematics Education, Hobart, Tasmania*, 13–18 July, Vol. 1, pp. 3–18). Hobart, Australia: PME Program Committee.

English, L. D., & Watson, J. M. (2015). Statistical literacy in the elementary school: Opportunities for problem posing. In F. Singer, N. Ellerton, & J. Cai (Eds.), *Problem posing: From research to effective practice* (pp. 241–256). Dordrecht, The Netherlands: Springer.

Esfandiari, M., Sorenson, K., Zes, D., & Nichols, K. (2014). Enhancing statistical literacy and thinking through analysis of scientific journal articles. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13–18)*. Voorburg, The Netherlands: International Statistical Institute.

Fielding-Wells, J. (2010). Linking problems, conclusions and evidence: Primary students' early experiences of planning statistical investigations. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society (Proceedings of the 8th International Conference on the Teaching of Statistics, Ljubljana, Slovenia, July 11–16)*. Voorburg, The Netherlands: International Statistical Institute.

Fielding-Wells, J., & Makar, K. (2012). Developing primary students' argumentation skills in inquiry-based mathematics classrooms. In K. T. Jan van Aalst, M. J. Jacobson, & P. Reimann (Eds.), *The Future of learning: Proceedings of the 10th International Conference of the Learning Sciences [ICLS 2012]—Volume 2 Short Papers, Symposia, and Abstracts* (pp. 149–153). Sydney, Australia: International Society of the Learning Sciences.

Fielding-Wells, J., & Makar, K. (2015). Inferring to a model: Using inquiry-based argumentation to challenge young children's expectations of equally likely outcomes. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 1–27). Minneapolis, MN: Catalyst Press.

Finzer, W. (2012). *Fathom dynamic data software [computer software, Version 2.13]*. Emeryville, CA: Key Curriculum Press and KCP Technologies.

Finzer, W., & Parvate, V. (2008). *Who will teach them about data?* Paper presented at the International Conference on Mathematics Education, Monterrey, Mexico.

Fitzallen, N. (2012). Interpreting graphs: Students developing an understanding of covariation. In J. Dindyal, L. P. Cheng, & S. F. Ng (Eds.), *Mathematics education: Expanding horizons* (*Proceedings of the 35th annual conference of the Mathematics Education Research Group of Australasia*, Vol. 1, pp. 290–297). Singapore: MERGA.

Fitzallen, N. (2013). Characterising students' strategies with *TinkerPlots*. *Technology Innovations in Statistics Education, 7*(1), Article 2.

Fitzallen, N., Watson, J., & English, E. (2015). Assessing a statistical inquiry. In K. Beswick, T. Muir, & J. Wells, (Eds.), *Climbing mountains, building bridges* (*Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education, July 13–18*, Vol. 2, pp. 305–312). Hobart, Australia: PME Program Committee.

Forster, M., & Wild, C. J. (2010). Writing about findings: Integrating teaching and assessment. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 87–102). Chichester, UK: Wiley.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A preK-12 curriculum framework*. Alexandria, VA: American Statistical Association.

Friel, S., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124–158.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–51.

Gal, I. (2005). Towards "Probability Literacy" for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 39–63). New York: Springer.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *BMJ: British Medical Journal, 327*(7417), 741–744.

Gil, E., & Ben-Zvi, D. (2010). Emergence of reasoning about sampling among young students in the context of informal inferential reasoning. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society (Proceedings of the 8th International Conference on the Teaching of Statistics, Ljubljana, Slovenia, July 11–16)*. Voorburg, The Netherlands: International Statistical Institute.

Goodchild, S. (1988). School pupils' understanding of average. *Teaching Statistics, 10*(3), 77–81.

Hammerman, J. (2009). *Exploring large scientific data sets as an entrée to statistical ideas in secondary schools*. Paper presented at the IASE Satellite conference: Next steps in statistics education, Durban, South Africa.

Holmes, P. (1980). *Teaching statistics 11–16*. Slough, UK: Schools Council Publications and W. Foulsham.

Hovermill, J., Beaudrie, B., & Boschmans, B. (2014). Statistical literacy requirements for teachers. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13–18)*. Voorburg, The Netherlands: International Statistical Institute.

Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics in the Middle School, 5*(4), 240–263.

Konold, C. (2007). Designing a data analysis tool for learners. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 267–291). New York: Lawrence Erlbaum.

Konold, C., Finzer, W., Kreetong, K., & Gaston, R. (2014, April). *Modeling as a core component of structuring data*. Paper presented at the annual meeting of the Research Conference of the National Council of Teachers of Mathematics, New Orleans.

Konold, C., & Harradine, A. (2014). Contexts for highlighting signal and noise. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Using tools for learning mathematics and statistics* (pp. 237–250). Heidelberg: Springer Spektrum.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

Konold, C., & Miller, C. D. (2015). *TinkerPlots: Dynamic data exploration [Computer software, Version 2.3]*. Adelaide, SA: Learn Troop.

Lane, D. M., & Peres, S. C. (2006). Interactive simulations in the teaching of statistics: Promise and pitfalls. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education (Proceedings of the 7th International Conference on the Teaching of Statistics, Salvador,* *Bahai, Brazil, July 2–7)*. Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute.

Lavigne, N. C., & Lajoie, S. P. (2007). Statistical reasoning of middle school children engaging in survey inquiry. *Contemporary Educational Psychology, 32*(4), 630–666.

Leavy, A. M., Friel, S. N., & Mamer, J. D. (2009). It's a fird! Can you compute a median of categorical data? *Mathematics Teaching in the Middle School, 14*(6), 344–351.

Lehrer, R., Kim, M., & Jones, R. S. (2011). Developing conceptions of statistics by designing measures of distribution. *ZDM Mathematics Education, 43*(5), 723–736.

Lehrer, R., Kim, M.-J., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modeling variability. *International Journal of Computers for Mathematical Learning, 12*(3), 195–222.

MacFeely, S., & MacCuirc, E. (2014). More ways to heaven than one: Improving statistical literacy in Ireland. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, AZ, July 13–18)*. Voorburg, The Netherlands: International Statistical Institute.

Madden, S. R. (2011). Statistically, technologically, and contextually provocative tasks: Supporting teachers' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1–2), 109–131.

Madden, S. R. (2013). Supporting teachers' instrumental genesis with dynamic mathematical software. In D. Polly (Ed.), *Common core mathematics standards and implementing digital technologies* (pp. 295–318). Hershey, PA: IGI Global.

Makar, K. (2010). Teaching primary teachers to teach statistical inquiry: The uniqueness of initial experiences. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society (Proceedings of the 8th International Conference on the Teaching of Statistics, Ljubljana, Slovenia, July 11–16)*. Voorburg, The Netherlands: International Statistical Institute.

Makar, K. (2014). Young children's explorations of average through informal inferential reasoning. *Educational Studies in Mathematics, 86*(1), 61–78.

Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning, 18*(1), 1–24.

Makar, K., Bakker, A., & Ben-Zvi, D. (2015). Scaffolding norms of argumentation-based inquiry in a primary mathematics classroom. *ZDM Mathematics Education, 47*(7), 1107–1120.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*, 152–173.

Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 1–4.

Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 347–358). Dordrecht, The Netherlands: Springer.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.

Manor, H., & Ben-Zvi, D. (2015). Students' articulations of uncertainty in informally exploring sampling distributions. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 57–94). Minneapolis, MN: Catalyst Press.

Meletiou-Mavrotheris, M., & Paparistodemou, E. (2015). Developing students' reasoning about samples and sampling in the context of informal inferences. *Educational Studies in Mathematics, 88*(3), 385–404.

Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education, 10*(1), Article 4.

Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, NZ: Author. Retrieved from http://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum

Ministry of Education. (2009). *The New Zealand curriculum: Mathematics standards for years 1–8*. Wellington, NZ: Author. Retrieved from http://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum/Learning-areas/Mathematics-and-statistics

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20–39.
Monk, S. (2003). Representation in school mathematics: Learning to graph and graphing to learn. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 250–262). Reston, VA: The National Council of Teachers of Mathematics.
Moore, D. S. (1990). Uncertainty. In L. S. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.
Moore, D. S., & McCabe, G. P. (1989). *Introduction to the practice of statistics*. New York: W. H. Freeman.
Moore, D. S., McCabe, G. P., & Craig, B. A. (2014). *Introduction to the practice of statistics* (8th ed.). New York: W. H. Freeman.
Moreno, J. L. (2002). Toward a statistically literacy citizenry: What statistics everyone should know. In B. Phillips (Ed.), *Developing a statistically literate society (Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town, South Africa, July 7–12)*. Voorburg, The Netherlands: International Statistical Institute.
Murray, S., & Gal, I. (2002). Preparing for diversity in statistics literacy: Institutional and educational implications. In B. Phillips (Ed.), *Developing a statistically literate society (Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town, South Africa, July 7–12)*. Voorburg, The Netherlands: International Statistical Institute.
National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
North, D., Zewotir, T., & Gal, I. (2014). Developing statistical literacy amongst in-service teachers through a collaborative project. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, AZ, July 13–18)*. Voorburg, The Netherlands: International Statistical Institute.
Osana, H. P., Leath, E. P., & Thompson, S. E. (2004). Improving evidential argumentation through statistical sampling: evaluating the effects of a classroom intervention for at-risk 7th-graders. *The Journal of Mathematical Behavior, 23*(3), 351–370.
Pea, R. (1985). Beyond amplification: Using the computer to reorganize mental functioning. *Educational Psychologist, 20*(4), 167–182.
Peck, R., Gould, R., & Miller, S. (2013). *Developing essential understanding of statistics for teaching mathematics in grades 9–12*. Reston, VA: National Council of Teachers of Mathematics.
Pfannkuch, M. (2005). Thinking tools and variation. *Statistics Education Research Journal, 4*(1), 83–91.
Pfannkuch, M., Arnold, P., & Wild, C. (2015). What I see is not quite the way it is: Students' emergent reasoning about sampling variability? *Educational Studies in Mathematics, 88*(3), 343–360.
Phillips, A. M. (2002). DNA "fingerprints" and their statistical analysis in human populations. In B. Phillips (Ed.), *Developing a statistically literate society (Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town, South Africa, July 7–12)*. Voorburg, The Netherlands: International Statistical Institute.
Pollatsek, A., Lima, S., & Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics, 12*(2), 191–204.
Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical influence. *Statistics Education Research Journal, 7*(2), 107–129.
Rao, C. R. (1975). Teaching of statistics at the secondary level: An interdisciplinary approach. *International Journal of Mathematical Education in Science and Technology, 6*, 151–162.
Reaburn, R. (2014). Introductory statistics course tertiary students' understanding of *p*-values. *Statistics Education Research Journal, 13*(1), 53–65.
Rubin, A., Bruce, B., & Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *School and general issues (Proceedings of the 3rd International*

Conference on the Teaching of Statistics, Dunedin, New Zealand, August 19–24). Voorburg, The Netherlands: International Statistical Institute.

Russell, S. J., & Mokros, J. (1990). What's typical? Children's and teachers' ideas about average. In D. Vere-Jones (Ed.), *School and general issues (Proceedings of the 3rd International Conference on the Teaching of Statistics, Dunedin, New Zealand, August 19–24)*. Voorburg, The Netherlands: International Statistical Institute.

Saldanha, L., & McAllister, M. (2014). Using re-sampling and sampling variability in an applied context as a basis for making statistical inferences with confidence. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13–18)*. Voorburg, The Netherlands: International Statistical Institute.

Santos, R., & Ponte, J. P. (2014). Learning and teaching statistical investigations: A case study of a prospective teacher. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13–18)*. Voorburg, The Netherlands: International Statistical Institute.

Schield, M. (2002). Three kinds of statistical literacy: What should we teach? In B. Phillips (Ed.), *Developing a statistically literate society (Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town, South Africa, July 7–12)*. Voorburg, The Netherlands: International Statistical Institute.

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph & K. Carr (Eds.), *People in mathematics education* (Proceedings of the 20th annual conference of the Mathematics Education Research Group of Australasia, Vol. 1, pp. 6–22), Waikato, New Zealand: MERGA.

Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In G. Burrill & P. Elliott (Eds.), *Thinking and reasoning with data and chance* (pp. 77–98). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M., & Pfannkuch, M. (2002). How faithful is old faithful? Statistical thinking: A story of variation and prediction. *Mathematics Teacher, 95*(4), 252–259.

Sproesser, U., Kuntze, S., & Engel, J. (2014). A multilevel perspective on factors influencing students' statistical literacy. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13–18)*. Voorburg, The Netherlands: International Statistical Institute.

Strauss, S., & Bichler, E. (1988). The development of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education, 19*(1), 64–80.

Till, C. (2014). Risk literacy: First steps in primary school. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13–18)*. Voorburg, The Netherlands: International Statistical Institute.

Toulmin, S., Rieke, R. D., & Janik, A. (1984). *An introduction to reasoning*. New York: McMillan.

Trouche, L. (2005). Instrumental genesis, individual and social aspects. In D. Guin, K. Ruthven, & L. Trouche (Eds.), *The didactical challenge of symbolic calculators: Turning a computational device into a mathematical instrument* (pp. 197–230). New York: Springer Science and Business Media.

Utts, J. (2002). What educated citizens should know about statistics and probability. In B. Phillips (Ed.), *Developing a statistically literate society (Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town, South Africa, July 7–12)*. Voorburg, The Netherlands: International Statistical Institute.

Watson, J. M. (2005). Variation and expectation as foundations for the chance and data curriculum. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce & A. Roche (Eds.), *Building connections: Theory, research and practice* (*Proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia, Melbourne*, pp. 35–42). Sydney: MERGA.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.

Watson, J. M. (2009). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal, 8*(1), 32–61.

Watson, J. M. (2016). Linking science and statistics: Curriculum expectations in three countries. *International Journal of Science and Mathematics Education, 15*(6), 1057–1073.

Watson, J. M., & Callingham, R. A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3–46.

Watson, J., Chick, H., & Callingham, R. (2014). Average: The juxtaposition of procedure and context. *Mathematics Education Research Journal, 26*(3), 477–502.

Watson, J., & Donne, J. (2009). TinkerPlots as a resource tool to explore student understanding. *Technology Innovations in Statistics Education, 3*(1), Article 1.

Watson, J., & English, L. (2015). Introducing the practice of statistics: Are we environmentally friendly? *Mathematics Education Research Journal, 27*(4), 585–613.

Watson, J., & Fitzallen, N. (2016). Statistical software and mathematics education: Affordances for learning. In L. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (3rd ed., pp. 563–594). New York: Taylor and Francis.

Watson, J. M., Fitzallen, N., & Carter, P. (2013). *Top drawer teachers: Statistics*. Adelaide, Australia: Australian Association of Mathematics Teachers and Education Services Australia. Retrieved from http://topdrawer.aamt.edu.au/Statistics

Watson, J. M., Fitzallen, N. E., Wilson, K. G., & Creed, J. F. (2008). The representational value of hats. *Mathematics Teaching in the Middle School, 14*(1), 4–10.

Watson, J., & Kelly, B. A. (2005). Cognition and instruction: Reasoning about bias in sampling. *Mathematics Education Research Journal, 17*(1), 25–57.

Watson, J. M., & Moritz, J. B. (1999). The development of the concept of average. *Focus on Learning Problems in Mathematics, 21*(4), 15–39.

Watson, J. M., & Moritz, J. B. (2000a). Development of understanding of sampling for statistical literacy. *The Journal of Mathematical Behavior, 19*(1), 109–136.

Watson, J. M., & Moritz, J. B. (2000b). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2*(1 & 2), 11–50.

Whitaker, D., Foti, S., & Jacobbe, T. (2015). The levels of conceptual understanding in statistics (LOCUS) project: Results of the pilot study. *Numeracy: Advancing Education in Quantitative Literacy, 8*(2), Article 3.

Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal, 5*(2), 10–26.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–248.

Zakaria, E., & Salleh, F. (2012). Teachers' creativity in posing statistical problems from discrete data. *Creative Education, 3*(8), 1380–1383.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education, 16*(2), Article 8.

Chapter 5
Reasoning About Data

Rolf Biehler, Daniel Frischemeier, Chris Reading, and J. Michael Shaughnessy

Abstract Many decisions in politics, economics, and society are based on data and statistics. In order to participate as a responsible citizen, it is essential to have a solid grounding in reasoning about data. Reasoning about data is a fundamental human activity; its components can be found in nearly every profession and in most school curricula in the world. This chapter reviews past and recent research on reasoning about data across all ages of learners from primary school to adults. Specifically in this chapter, the term reasoning about data is defined, the implementation of reasoning about data in the curricula of different countries is investigated, and research studies of learner reasoning about distribution, variation, comparing groups, and association, which are fundamental concepts when reasoning about data, are reviewed. The research review presented includes references to existing frameworks and taxonomies that can assess learner reasoning in regard to these concepts and discusses the influence of digital tools to enhance learner statistical reasoning. Finally, some insights for future directions in research about reasoning about data are provided.

Keywords Statistical reasoning • SOLO • Data • Distribution • Variation • Comparing groups • Association

R. Biehler (✉) • D. Frischemeier
Institute of Mathematics, University of Paderborn, Paderborn, Germany
e-mail: biehler@math.upb.de; dafr@math.upb.de

C. Reading
SiMERR National Research Centre, University of New England, Armidale, NSW, Australia
e-mail: creading@une.edu.au

J.M. Shaughnessy
The Fariborz Maseeh Department of Mathematics and Statistics,
Portland State University, Portland, OR, USA
e-mail: mikesh@pdx.edu

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_5

5.1 Introduction

"Data really powers everything that we do!" (Jeffrey Weiner, chief executive of LinkedIn)

The purpose of this chapter is to provide an overview of research on reasoning about data across all age levels. Four general phases can be identified in a data analysis process (e.g., Graham, 1987; Kader & Perry, 1994; Wild & Pfannkuch, 1999): (1) pose a statistical question and generate a hypothesis, (2) collect data, (3) analyze data, and (4) interpret the results and communicate conclusions. This chapter emphasizes the last two phases, analyzing data and interpreting results, focusing on research that looks at students and learners' reasoning about *variation*, *distribution*, *group comparisons*, and *associations between variables*.

The chapter highlights frameworks and taxonomies that can help to understand a learner's reasoning, as well as provide insight into the development of research in a specific domain. Since there is much research based on the Structure of Observed Learning Outcome (SOLO) Taxonomy (Biggs & Collis, 1991), the discussion also highlights aspects relevant to SOLO. Furthermore, different sections of this chapter outline possibilities and potentials to use technology to support a better and more sophisticated reasoning about data.

This chapter provides a historical perspective on the evolution of research in student reasoning about data. Earlier work on conceptual frameworks describing student reasoning about data led to more focused research on concepts such as variation and distribution. The chapter is divided into seven sections. Section 5.2 considers general elements in regard to reasoning about data and presents several frameworks on reasoning about data. Section 5.3, on "reasoning about variation," includes reasoning in contexts and provides an overview of research across all age levels with a special focus on developmental hierarchies of reasoning. Section 5.4, on "reasoning about distribution," includes perspectives on the concept of distribution and research on student reasoning about distribution, including developmental models. Section 5.5, on "reasoning when comparing groups," distinguishes several facets of group comparisons and reviews research on comparing groups with students and pre- and in-service teachers. Section 5.6, on "reasoning about associations between variables," focuses on studies regarding learners' reasoning about associations in contingency tables and between numerical variables. The final section concludes the chapter with a discussion and an outline for further research.

5.2 Reasoning About Data

This section begins with a discussion of some examples of positions on reasoning about data put forth in four national curriculum recommendations for statistics education (Australia, Germany, the United States of America, and New Zealand). The section ends with some broad-based perspectives from research on frameworks for researching student reasoning about data. Thus, the section provides a foundation for the remaining sections in the chapter that delve in more detail into research on students' reasoning about some major concepts in data: variation, distributions, comparing groups, and association.

5.2.1 Some National Recommendations on Reasoning About Data

Over the past 25 years, the inclusion of statistics and probability in national school curricula has continued to grow in prominence. Many nations now provide curricular documents that list, and in some cases discuss, the topics or concepts that should be included in the statistical education of elementary and secondary school students at various grade levels. For example, in Australia, guidelines for Foundation (Kindergarten) to Year 7 (ages 4–12) concentrate primarily on the construction and comparison of a range of data displays, while guidelines for Year 8 and Year 9 have students explore variation of means and proportions in random samples, look at shapes of distribution, learn about sampling distributions, and compare distributions using measures of center and variation. In later school years, the Australian curriculum (n.d.) concentrates more on probability models, introduces random variables, and begins statistical inference (ACARA, n.d.). Thus, in Australia a trajectory for students' statistical education has been laid out for every year they are in elementary and secondary schools.

In Germany, guidelines for Grades 1–4 similarly concentrate on collecting and representing data as well as on how to use representations of data (Hasemann & Mirwald, 2012). Grades 5–10 concentrate on data analysis, interpretation, and developing arguments based on data (KMK, 2004). The education competency standards for mathematics in Grades 11–13 include inference from samples to populations, simulations, probability distributions, and hypothesis testing (KMK, 2012).

The recommendations for statistical education of students in Germany and Australia appear to be quite similar. Early grades concentrate on exploring and representing data. The middle grades ask students to carry out more detailed analysis of data and to begin to make arguments based on reasoning about data (this appears to start earlier in Germany, Grades 5–10, than in Australia, Grades 8 and 9). Upper secondary school students in both countries begin to make inference from samples, explore sampling distributions, and begin to do some hypothesis testing.

For more than 20 years, detailed recommendations for the statistical education of Grades pre-K–12 students have been published in the standards documents of the National Council of Teachers of Mathematics (2000; NCTM, 1989; Shaughnessy, Chance, & Kranendonk, 2009). The recommendations for data analysis and probability in the *Principles and Standards* document (NCTM, 2000) are elaborated under four main clusters for students in four grade bands (Grades pre-K–2, 3–5, 6–8, and 9–12). These four clusters state that instructional programs in data analysis should enable students to: (i) formulate questions that can be addressed with data, (ii) select and use appropriate statistical methods to analyze data, (iii) develop and evaluate inferences and predictions based on data, and (iv) understand and apply basic probability concepts. In recent years, the Common Core State Standards for Mathematics (Common Core State Standards Initiative, 2015) have been adopted by a number of states in the United States of America, and they include clusters of statistics and probability standards for Grades 6–11. The middle school clusters start with understanding variability and summarizing and describing distributions of data (Grade 6), continue with inference from a sample to a population and informal comparative inferences about two populations (Grade 7), and conclude with association and bivariate data (Grade 8). The secondary standards interweave probability models and approaches to statistical inference, including hypothesis testing via simulations and bootstrapping. Unlike the NCTM recommendations, the Common Core Standards do not include recommendations for reasoning about data in Grades 3–5. However, the use of displays for measurement data is a topic that is included in the elementary years in the Common Core.

National curricula documents do not always include sufficient detail on the importance of *reasoning about data*, or what such reasoning about data would involve. An exception can be found in New Zealand where the statistics curriculum has made recommendations for the inclusion of the *processes* students should experience in conducting investigations using the statistical investigation cycle (Ministry of Education, 2007). The German standards also refer to process skills related to data analysis (KMK, 2004). The NCTM standards for statistics and data analysis were available to many other countries as they established their own recommendations for statistics education and were likely to have been an influence on the development of statistics standards in Australia, Germany, and New Zealand, among other countries.

Some professional statistics organizations such as the American Statistical Association (ASA) have published detailed recommendations for students to engage in the *process* of reasoning about data. The pre-K–12 GAISE[1] report of the ASA contains recommendations for data analysis for elementary, middle, and secondary level students (Franklin et al., 2007). Reasoning about data in the pre-K–12 GAISE report goes far beyond just a bulleted list of topics that one may find in many national curriculum documents. The pre-K–12 GAISE report recommends that students should" develop strategies for producing, interpreting, and analyzing data to help answer questions of interest. In general, students should be able to formulate questions that can be answered with data; devise a reasonable plan for collecting appropriate data through observation, sampling, or experimentation; draw conclusions using data to support these conclusions; and understand the role random variation plays in the inference process" (Franklin et al., 2007, p. 61). In the pre-K–12 GAISE report, the entire process of reasoning about data is anchored in the importance of posing a statistical question in the first place.

[1] There is also a GAISE report for the statistical education of college students.

The recommendations for pre-K–12 statistics education contained in these national documents and statistics organizations have undoubtedly led to the development of a number of curriculum innovations and the development of statistical materials for teaching over the past three decades. However, statistics curriculum materials are not a topic that is within the purview of this chapter. Given these many recommendations about what students should experience in their statistics education, the interesting question for research is: How *do* students reason about data when given opportunities to do so? The next section considers frameworks that may be useful to researchers who are interested in exploring students' statistical reasoning about data.

5.2.2 Frameworks on Reasoning About Data

Early twenty-first century statistics education research found researchers striving to build conceptual frameworks to interpret students' statistical reasoning. Some such frameworks for statistical reasoning were based on hypothesized statistical processes. For example, Jones, Langrall, Mooney, and Thornton (2004) analyzed the empirical responses of middle school and elementary school students to statistical tasks and represented the identified diverse aspects of reasoning about data based on four key statistical processes: describing data, organizing data, representing data, and analyzing and interpreting data. Their investigation led them to suggest descriptors for four levels of increasingly sophisticated student reasoning about data: idiosyncratic, transitional, quantitative, and analytical. Their work suggested that future research is needed to analyze responses to similar tasks from older learners to see whether the four levels of reasoning about data (idiosyncratic, transitional, quantitative, and analytical) can be validated and extended to reasoning by secondary and tertiary students. For a more detailed description of this model and an analysis of early models of development in statistical reasoning, see Jones et al. (2004).

Conceptual frameworks that focus on reasoning about specific statistical concepts (e.g., centers, variability, distribution) or particular statistical processes (e.g., organizing data, reducing data) may themselves be anchored in more general theoretical models of reasoning. In particular, the Structure of Observed Learning Outcome (SOLO) Taxonomy (Biggs & Collis, 1991) has been utilized by a number of researchers to analyze and characterize levels of increasing complexity of understanding demonstrated in student responses to statistical tasks. This taxonomy is a neo-Piagetian model of cognitive development created to analyze the complexity of student responses to tasks and was developed by Biggs and Collis (1991) as a general model for evaluating learning in any context or environment. SOLO identifies five modes of functioning (sensorimotor, ikonic, concrete-symbolic, formal, and post-formal), with a number of achievements identifiable within each of these modes. The two modes most relevant to school-aged student reasoning are the ikonic mode (making use of imaging and imagination) and the more cognitively complex concrete-symbolic mode (operating with second-order symbol systems such as written language). The formal mode (reasoning that generates generalizations that both incorporate and transcend particular situations) is often not observed in responses until students are learning at late secondary or tertiary levels. Although these modes are similar to Piagetian stages, there is an important difference. Earlier modes in the SOLO Taxonomy are viewed as being used to support growth in later modes rather than as being replaced by later modes.

Within each of these modes, a series of levels of increasing cognitive development has been identified. The three levels most commonly reported in research are: unistructural (U) responses, which focus on one relevant aspect of the task; multistructural (M) responses, which focus on several relevant aspects separately; and relational (R) responses, which focus on several relevant aspects in which interrelationships are identified. These three levels form a cycle of cognitive growth, from U, through M, to R that occurs within each mode. For example, Reading and Reid (2006) studied students' reasoning about distribution based on a variety of in-class tasks and described a hierarchy for coding responses, within the concrete-symbolic mode, based on the aspect *key element of distribution*. A U level response focused on one element of distribution (e.g., center, spread, density, skewness, or outliers), an M level response focused on multiple elements of distribution without linking them, and an R level response integrated all available elements of distribution into the response.

Overall, SOLO has proven to be a useful research tool for statistics educators since it is designed to assess responses to open-ended complex tasks that can elicit a hierarchy of student reasoning. As researchers have focused more on specific statistical domains or concepts, they have been able to investigate reasoning in more depth, which at times has resulted in the need to describe more than one U-M-R learning cycle (e.g., Reading & Reid, 2006; Watson & Moritz, 1999, 2000). Some researchers believe that when there are two such cycles, the first cycle of learning relates to the development of the concept, while the second cycle deals with the consolidation and application of the concept (Jones et al., 2004).

However, researchers have discovered that generalized descriptions (such as those in SOLO) do not always provide enough detail to represent all the important aspects of reasoning about data. The result has been the development of a number of domain-specific and concept-specific reasoning frameworks relevant to reasoning about data. In a review of the research literature on statistics learning and reasoning, Shaughnessy (2007) noted that researchers were beginning to build developmental conceptual frameworks of student reasoning about centers (Watson & Moritz, 2000), graphs (Friel, Curcio, & Bright, 2001), and variability and distribution (Bakker & Gravemeijer, 2004; Saldanha & Thompson, 2003; Shaughnessy, Ciancetta, & Canada, 2004). Some of the developmental frameworks are hierarchical in nature, such as the interpretive conceptual frameworks reported by Ben-Zvi (2004) and Noll and Shaughnessy (2012). These developmental frameworks suggest that students may need to progress through particular stages of reasoning as they gain more experience with statistics.

5.2.3 Summary and Preview

This section provided a brief overview of what is meant by statistical reasoning about data from the perspective of national curriculum documents and some statistics and mathematics education organizations. This was followed by discussion of some early fundamental frameworks for researching student reasoning about data. Such research has since moved on from considering overall reasoning about data to a focus on reasoning about particular aspects or characteristics of data. The discussion in the remaining sections focuses on research that involves reasoning about data through the lens of particular statistical concepts: variation, distribution, group comparison, and association.

5.3 Reasoning About Variation

This section considers the fundamental role of reasoning about variation, then focuses on frameworks developed to assist researchers and teachers to delve into reasoning about variation as they support learners. First, the essence of variability and variation are considered. Second, the breadth of contexts in which variation is studied is explained. Third, the initial growth of developmental hierarchies informing work on reasoning about variation is investigated. Fourth, those hierarchies which are SOLO-informed are synthesized. Fifth, the research into reasoning exhibited by teachers and preservice teachers is examined. Finally, conceptions of variation as evidenced in reasoning are elaborated.

5.3.1 Exploring Reasoning About Variation

The notion of variation is fundamental to statistical thinking because analysis seeks to "explain systematic effects behind the random variability of individuals and measurements" (Pfannkuch & Wild, 2004, p. 38). While some authors use the term variability interchangeably with the term variation (see, e.g., Peters, 2011), Reading and Shaughnessy (2004) explained that variability is an observable characteristic of an entity while variation is associated with measuring that characteristic. Research that informs this section has been reported using the term variability or variation as used by the respective researchers, even though this may not be consistent with the distinction made above between observable characteristic (variability) and the measurement of that characteristic (variation).

The focus of this section is to examine reasoning about variation, that is, reasoning used to deal with situations that exhibit change, i.e., variability. Four main sources of variability are recognized for statistics education settings: measurement, natural, induced, and sampling (Franklin et al., 2007). Measurement variability occurs when repeated measurements taken on the same individual entity vary, perhaps due to the reliability of the measuring device or due to actual changes in what is being measured. Natural variability occurs because individuals are inherently different and so measurement of the same quantity over different individuals is likely to vary. Induced variability occurs when other factors are introduced that change conditions and thus necessarily change measurement. Sampling variability occurs when more than one sample is drawn and the measures calculated from the samples vary. Being able to reason about variation, when the variability is from these diverse sources, provides a strong basis for statistical thinking.

The study of reasoning about variation is often informed by research into "consideration of variation," which is one of the fundamental types of statistical thinking identified by Pfannkuch and Wild (2004). Consideration of variation was proposed to have four components: (1) noticing and acknowledging variation; (2) measuring and modeling variation for the purpose of prediction, explanation, or control; (3) explaining and dealing with variation; and (4) developing investigative strategies in relation to variation. This list was later expanded (Reading & Shaughnessy, 2004) to include two more components: (5) describing variation and (6) representing variation, which are particularly important for school students in their early consideration of variation.

Research into "understanding variation" has also informed research on reasoning about variation. After reviewing research to date, Garfield and Ben-Zvi (2005) proposed a theoretical framework of seven key facets of "understanding variation." A synthesis of this framework and contributions from a number of key researchers resulted in Reading and Reid (2010) proposing a framework of nine facets/components of variation: (1) developing intuitive ideas of variability, (2) describing and representing variability, (3) using variability to make comparisons, (4) recognizing variability in special types of distributions, (5) identifying patterns of variability in fitting models, (6) using variability to predict random samples or outcomes, (7) considering variability as part of statistical thinking, (8) recognizing sources of variation, and (9) resolving expectations with observed variation. This nine-facet framework expands the view of understanding variation, especially in relation to dealing with variation in situations involving prediction, explanation, and control. Such depth in the framework potentially provides a better focus for teachers as they plan learning experiences to assist students to reason about variation.

5.3.2 Reasoning About Variation in Context

The very nature of variation necessitates the study of reasoning about variation in context. The context is necessary to provide meaning while reasoning (Franklin et al., 2007). When investigating reasoning, contexts can be naturally occurring or artificial (nonnaturally occurring) experiences. Artificial experiences, such as using chance devices, have been found to be helpful for investigating reasoning because there are fewer sources of variation to consider, even though naturally occurring situations might motivate an understanding of variation (Watson & Kelly, 2004a).

Reasoning about variation has been studied under a variety of contexts/situations. Some studies have asked students to reason about different types of representations (e.g., graphs, tables, models; Pfannkuch, 2005) or used technology-based simulations (Lehrer, Kim, & Schauble, 2007). A variety of activities have also been used: describing distributions (Reading & Reid, 2006), comparing within and/or between distributions (Ben-Zvi, 2004; Makar & Confrey, 2005; Pfannkuch, 2005; Reid & Reading, 2008), modeling (Lehrer & Schauble, 2004), predicting outcomes (Mooney, Duni, VanMeenen, & Langrall, 2014; Watson, Kelly, Callingham, & Shaughnessy, 2003), or drawing conclusions (Peters, 2011).

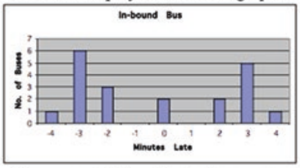
There are three key contexts in which reasoning about variation has been researched: (1) chance, (2) data and graphs, and (3) sampling. Examples of research using chance contexts include investigating expectation of spinner trials (Canada, 2006; Watson & Kelly, 2004a) and six-sided die rolls (Watson et al., 2003). Examples using data and graph contexts include describing weather data (Reading, 2004), summarizing and analyzing bird egg data (Reid & Reading, 2008), comparing lengths of surnames (Ben-Zvi, 2004), and comparing student grades (Makar & Confrey, 2005). Examples that used sampling contexts include predicting outcomes selecting from a candy bowl (Mooney et al., 2014; Reading & Shaughnessy, 2004), modeling plant growth (Lehrer & Schauble, 2004), and expressing expectation about student weights (Watson & Kelly, 2006). All three key contexts were used by Reading and Reid (2006) when they designed a questionnaire to investigate tertiary students' reasoning about variation before and after an introductory statistics course (see Fig. 5.1). The first question was designed to determine whether the students knew about variability, while questions 2, 3, and 4 were designed to investigate their reasoning about variation in the contexts of data and graphs, chance, and sampling, respectively.

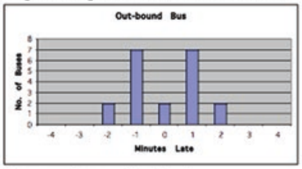
Much research into reasoning about variation combined several contexts to create a richer view of the reasoning. Some researchers (e.g., Reid & Reading, 2006, 2008) combined items from various contexts into one instrument, while others (e.g., Watson et al., 2003; Watson, Callingham, & Kelly, 2007) combined items from all three key contexts to create, by statistical analysis (Rasch modeling), one construct for reasoning about variation. Peters (2011) provided rich insight into students' reasoning by combining contexts across three different perspectives (design, data-centric, and modeling) to code responses from a variety of tasks. For example, Peters found that insight into how teachers' reason about controlling variation was better derived from small sample situations when such situations were placed in contexts.

A variety of studies with a focus on the role of context when students are developing their reasoning in an informal statistical inference situation, summarized in Makar and Ben-Zvi (2011), drew attention to the importance of context when studying reasoning but one serious warning was shared. Context knowledge, which can influence the way students interact with data, may actually cause students to look beyond the data and use their context knowledge to provide explanations of patterns

Question 1 - What does variability mean to you? Give a verbal explanation and/or an example.

Question 2 - Citizens in an outer suburb were concerned about the reliability of their bus service to the centre of the city. They monitored the in-bound and out-bound service of the buses at Bus Stop 33, and recorded the number of minutes late. Zero minutes late indicates the bus was on time while a negative number of minutes late indicates the bus was early. The data are displayed in the two graphs. Describe and compare the performances of the two bus routes.





*Question 3** - Every year in New Zealand approximately seven children are born with a limb missing. Last year the children born with this abnormality were located in New Zealand as shown on the map. In New Zealand, it is common knowledge that one-third of the population lives in the top region and one-sixth of the population in each of the other four regions. What do you think? ** Meletiou-Mavrotheris & Lee (2002)*

*Question 4*** - A bowl has 100 wrapped hard candies in it. 20 are yellow, 50 are red, and 30 are blue. They are well mixed up in the bowl. Jenny pulls out a handful of 10 candies whilst blindfolded, counts the number of reds, and tells her teacher. The teacher writes the number of red candies on a list. Then, Jenny puts the candies back into the bowl, and mixes them all up again. Five of Jenny's classmates, Jack, Julie, Jason, Jane and Jerry do the same thing. They each pick ten candies, count the reds, and the teacher writes down the number of reds. Then they put the candies back and mix them up again each time.

a. What do you think the teacher's list for the number of reds is likely to be? Explain why you chose those numbers.

b. If you were asked to choose a response to this question from the following list, circle the one that you would choose. Explain why you chose that one.
A) 5,9,7,6,8,7 B) 3,7,5,8,5,4 C) 5,5,5,5,5,4 D) 2,4,3,4,3,4 E) 3,0,9,2,8,5

c. All the students in Jenny's class watched the demonstration and wanted to take part. The teacher began the trial again, recording the results in a new list, allowing each student in the class of 40 to draw out 10 candies under the same controlled conditions. Describe a list that the teacher would have been likely to record. Explain why you described the list that way. *** Adapted from Reading & Shaugnessy (2004)*

Fig. 5.1 Questionnaire to investigate reasoning about variation (Reid & Reading, 2006, p. 2)

in or conclusions from the data. When working on group comparison tasks, middle school students openly admitted that data-based conclusions were not the only ones that could be drawn from the comparisons (Langrall, Nisbet, Mooney, & Jansem, 2011). In fact, some research has shown that if context knowledge is applied directly then context may become the only resource for providing such explanations. Such a tendency was identified in research when students, engaged in an inquiry-based project, had to deal with the conflict between what they expected during their investigation and what they were interpreting from the data (Makar, Bakker, & Ben-Zvi, 2011). Despite the potential to rely too heavily on it, these researchers included context knowledge as one of the important elements supporting informal inferential reasoning.

A similarly balanced approach should be taken to the place of context when students are reasoning about variation. While acknowledging that context is useful for situating such reasoning, care should be taken to encourage students to base their reasoning on the data provided, using their knowledge of the context to help make sense of their reasoning. This is in keeping with the general trend to engage students in tasks that involve a more complete statistical process, such as informal inferential (Makar & Ben-Zvi, 2011), and not just reasoning about variation in isolation. More use of experiences that enable modeling to understand phenomena is encouraged in order to ground reasoning within the complete statistical enquiry cycle (e.g., Lehrer et al., 2007; Pfannkuch, 2005). In particular, the GAISE Framework (Franklin et al., 2007) proposes that anticipating variability is critical to posing suitable questions to begin a statistical investigation, while acknowledging, accounting for, and allowing for variability become important in the later stages of an investigation.

5.3.3 Developmental Hierarchies of Reasoning

Developmental hierarchies of reasoning about the core concept of variation are needed to help teachers plan suitably structured learning sequences and relevant assessment of students' learning about statistical investigations. Developmental hierarchies were initially created to describe increasing sophistication in reasoning about variation. There are two important examples. First, Watson et al. (2003) described four levels: (1) prerequisites for variation, (2) partial recognition of variation, (3) applications of variation, and (4) critical aspects of variation. Second, Ben-Zvi (2004) classified reasoning into seven stages: (1) focus on irrelevant aspects, (2) describing variability, (3) forming a hypothesis to account for variability, (4) accounting for variability by comparing frequencies, (5) using measures of center and spread, (6) handling outlying values, and (7) distinguishing between within and between group variability.

Hierarchies developed by one research project cannot always be applied to another research project. After attempting to utilize pre-existing hierarchies, Slauson (2008) resorted to creating a new hierarchy in order to represent cognitive levels in sufficient detail. Similarly, Reid and Reading (2010) combined the Wild and Pfannkuch (1999) framework and the Reading and Shaughnessy (2004) Description Hierarchy to categorize reasoning about explained and unexplained variation, but needed to add an extra category (modeling and quantifying explained and unexplained variation) to suitably code student responses.

Increasing sophistication in such hierarchies is evident in the articulation of parallel aspects of developing reasoning about variation. Reading and Shaughnessy (2004) developed two multilevel parallel hierarchies: one for describing variation and the other for identifying the cause of variation. Similarly, Reid and Reading (2010) identified two developmental paths: one for modeling and quantifying explained and unexplained variation and the other for identifying and controlling causes of variation. More so, Peters (2011), in order to create a "robust" model for understanding variation, created three parallel hierarchies to represent three perspectives: design, data-centric, and modeling. Describing parallel aspects or perspectives of reasoning allows researchers to delve more deeply into interpreting student responses as they reason and allows teachers to plan more effective learning sequences to promote such reasoning.

Accurate articulation of level descriptors in developmental hierarchies is critical to identifying reasoning at a particular level and depends on both the quality and the variety of the task(s). For example, Reading and Reid (2006) initially described four levels of consideration of variation (no, weak, developing, strong) separately for each of four tasks. After analyzing responses to two further tasks, Reid and Reading (2008) improved the level descriptions by clearly stating differences between the levels in relation to how within-group variation and between-group variation were incorporated into the reasoning. Weak responses dealt with one type of variation or the other, developing responses dealt with both types of variation but as separate characteristics, and it was only the strong responses that incorporated both types of variation and explained how they were linked. This revised hierarchy has been utilized by other researchers (e.g., Mooney et al., 2014), who found that students recognized that variability should be expected but did not know how much variability to expect.

While the above has shown that a greater depth of hierarchical description of reasoning has generally been achieved when responses are given to tasks set in specific contexts, there may be other ways to further refine descriptions of reasoning. One effective approach is to challenge learners with cognitive conflict. Studies (e.g., Reading & Reid, 2007; Watson & Kelly, 2004a) reported that learners find it difficult to resolve such conflict but do share more detailed reasoning in their attempts. A related approach, resolving dilemmas, assisted teachers as learners to articulate their reasoning about variation (Peters, 2014).

As the variety of hierarchies of reasoning about variation continued to increase, a common basis for understanding the frameworks was needed to assist teachers as practitioners to more easily access the power of using the hierarchies to inform the planning of learning experiences and assessment of learning.

5.3.4 SOLO-Informed Hierarchies

The SOLO Taxonomy has become a popular theoretical framework for informing developmental hierarchies in reasoning generally and reasoning about variation in particular. Some researchers acknowledge that SOLO has informed their work but do not explain explicitly how their hierarchies relate to the SOLO framework. For example, Watson et al. (2003) explained that SOLO informed the starting point for their analysis but no specific SOLO terminology appeared in their explanation or hierarchy. For instance, their Level 4, critical reasoning about variation, requires a consolidation of concepts, and although resembling the SOLO relational level, the explicit connection is not explained. Further analyses by these researchers (Watson & Kelly, 2004a, 2006) described specific SOLO-coded levels from ikonic to relational, but these hierarchies focused on reasoning in the task itself (i.e., the context the variation is in) rather than specifically about variation.

One of the first researchers to provide detailed descriptions of SOLO levels for reasoning about variation was Reading (2004), whose depth of analysis clearly indicated that some responses dealt with qualitative descriptions of variation, one cycle of SOLO levels (U1-M1-R1), while other responses dealt with quantitative descriptions of variation, a second cycle of SOLO levels (U2-M2-R2). When more than one

SOLO cycle exists, it is critical that clear descriptors are provided to distinguish the step from one cycle to another (in the above case, the step is being able to provide a quantitative description of variation) as well as to distinguish one level from the next. A study by Watson et al. (2007) was the first large-scale study to use psychometric methods (Rasch modeling) to justify the break from one SOLO level to the next. The detailed levels in the resulting pathway for distinguishing development of variation and expectation clearly showed a first cycle where expectation and variation are addressed separately, and then a second cycle where the two concepts are linked.

SOLO-based research has also identified critical steps in reasoning about variation within the relational level of a SOLO hierarchy. Examples are linking explained and unexplained variation (Reid & Reading, 2010), linking expectation and variation (Watson & Kelly, 2006), and linking proportional reasoning and variation (Reid & Reading, 2006). This is consistent with other research studies that have not explicated hierarchies, e.g., going from seeing variation as differences to seeing variation as structured differences (Lehrer & Schauble, 2004), changing from nonstandard to standard terminology (Makar & Confrey, 2005), and moving from inventing measures to using formal statistical measures (Lehrer et al., 2007).

After critically reviewing earlier hierarchies, Peters (2011) developed a reasoning-focused hierarchy labeled "robust understanding of variation" that identified three perspectives on variation. The first perspective, design, "integrates acknowledgement and anticipation of variation in the design of quantitative studies." The second perspective, data-centric, "integrates processes of representing, measuring and describing variation in exploratory data analysis." The last perspective, modeling, "integrates reasoning to fit models to patterns of variability in data and statistics, judging goodness of fit, and transforming data to improve fit" (Peters, 2011, p. 53). The three perspectives are separate in the first SOLO cycle but become integrated in the second SOLO cycle. This hierarchy extends earlier hierarchies that are based in the concrete-symbolic mode into the formal mode. Peter's hierarchy provides very detailed descriptors for recognizing responses from all three perspectives, focusing on the relational requirement for reasoning within each cycle.

5.3.5 Adult Reasoning About Variation

Much of the research into reasoning about variation has involved school-aged students and first year tertiary students. Research to investigate higher levels of reasoning about variation may require studies that involve adults as respondents. Studies with adults have mostly involved teachers and have focused on observing them as learners in both data and chance contexts rather than investigating their understanding of variation (Sanchez, Borim da Silva, & Coutinho, 2011). There have been a few exceptions. Peters' (2011) in-depth research with competent mathematics/statistics teacher-leaders facilitated the refinement of levels in the SOLO formal mode. Later investigation of these teacher-leaders' understanding of variation, based around adult learning theories, highlighted the importance of disorienting dilemmas as triggers for the formation of robust conceptions of variation (Peters, 2014). Such experiences may be likened in value to reasoning experiences generating cognitive conflict when challenging learners to develop a better understanding of variation.

When investigating how teachers talked about variation while comparing distributions, Makar and Confrey (2005) detailed four types of terms teachers used to describe variation (spread, low-middle-high, modal clump, and distribution chunks). Their analysis showed the linking of nonstandard to standard terms used by the teachers to describe variation. The results from research conducted by Canada (2006) with preservice teachers in a probability setting produced an evolving framework that incorporated three aspects of reasoning: (1) expecting variation (describing what is expected and why), (2) displaying variation (producing, evaluating, and comparing graphs), and (3) interpreting variation (cause and effect). The framework provided a useful elaboration of the variety of conceptions of variation used by the preservice teachers and is consistent with, but not as detailed as, the Reading and Reid (2010) framework. This early work with teachers and preservice teachers suggested that when these adults were first developing their reasoning, they articulated their ideas in similar ways to younger learners.

5.3.6 Conceptions of Variation as Evidenced in Reasoning

The hierarchies described thus far provide insight into levels of reasoning observed in learner responses to items and/or tasks, but the development of such reasoning depends on a strong foundational conception of variation. So, what does this conception look like and how is this evidenced in such reasoning? After reviewing a decade of research on variability, Shaughnessy (2007) reported the following conceptions of variation: (1) variability in particular values including extremes and outliers, (2) variability as change over time, (3) variability as whole range (the spread of all possible values), (4) variability as the likely range of a sample, (5) variability as distance or difference from a fixed point, (6) variability as the sum of residuals, (7) variability as covariation or association, and (8) variability as distribution.

Specific examples of these conceptions have been found in research studies. For example, when analyzing explanations of variation, Reading and Shaughnessy (2004) found use of both middle values and extreme values (as per (1) above). Lehrer and Schauble (2004) found use of the middle 50% (semi-quartile range) by fifth graders to describe spread (as per (3) above). Lehrer et al. (2007) found fifth and sixth grade children could conceptualize measures as a composition of true value and chance error (as per (6) above). Watson and Kelly (2004b) found evidence of both the point expectation view (as per (5) above) and distributed expectation view of variability (as per (8) above).

The ability to reason at a certain level about variation is dependent on the maturity of the conception of variation. For example, the conception (8) above, the more advanced ability to explain *variability as distribution*, would be necessary before a response to a task could achieve the Reading and Reid (2010) "(4) recognizing variability in special types of distributions."

5.3.7 Reflections and Future Developments

An increasing focus on the problem-solving process in statistics education requires a better understanding of students' conceptions as they reason about variation. The GAISE Report (Franklin et al., 2007) explained that with maturation in the problem-solving process comes increased complexity in the role of variation. Developmental hierarchies of reasoning about variation provide an in-depth view of such reasoning and frameworks in which teachers can position optimal teaching and assessment of the reasoning. Within this progression, there will always be key hurdles that students need to achieve. For example, before being able to compare natural variation with induced variation (Franklin et al., 2007), students need to be able to distinguish between natural variation and induced variation and reason about each of them separately.

The nine key facets of understanding variation (Reading & Reid, 2010) provide a good starting point for developing a more informed view of the breadth of understanding needed to appreciate what may be necessary when reasoning about variation. There is an ever-increasing variety of hierarchies available to explain the detail of this reasoning and the more complex the situation involving variation, the more complex is the hierarchy needed to explain the reasoning. The context in which a specific task is situated is crucial to deciding which hierarchy is most relevant. If, however, there is no hierarchy to meet the needs of a specific context/situation, then adjustment of an existing hierarchy or development of a new hierarchy may be necessary. The SOLO Taxonomy provides researchers and teachers with a useful tool to underpin understanding of increasing sophistication in reasoning about variation in existing and newly developed hierarchies.

Statistics education teaching continues to increase in importance and to deliver relevant learning teachers need to improve their statistical knowledge, especially in relation to reasoning about variation. Teachers engaging in professional development activities focusing on collaboration and intellectual conversations have been shown to improve their understanding of variation and thus reasoning about variation (Peters, 2014). Such professional learning should be a starting point for teachers. As part of their practice, teachers need to make use of developmental hierarchies to inform their planning of learning sequences to take learners from less formal to more formal articulations of reasoning. This should then naturally lead to teachers making use of hierarchies to inform assessment of learning outcomes. Hierarchies with the SOLO Taxonomy as a theoretical basis have proven to be particularly useful for teachers when assessing learning.

Although, statistics education research continues to provide more hierarchies to explain reasoning, the second decade of the twenty-first century has seen more of a focus on reasoning about distribution than reasoning about variation. Researchers should consider revisiting research around reasoning about variation in light of the new knowledge being shared by research into reasoning about other concepts. For example, researcher focus on reasoning from samples and about sample means, including relevant variability, still acknowledges dependence on foundation statistical concepts, including variation, as fundamental to more advanced reasoning when dealing with data (Ben-Zvi, Bakker, & Makar, 2015).

Also, researchers should consider the possibility of utilizing a pre-existing hierarchy when relevant to the context, even though it may need to be adjusted to suit specific needs, as development of a new hierarchy is a time-consuming activity. Importantly, researchers should remember that the lack of use of formal conceptions and terminology of variation does not necessarily mean that the learner is not developing an appreciation of the concept. The importance of students developing the ability to view data from a variety of perspectives, as they work toward “data as aggregate” (Konold, Higgins, Russell, & Khalil, 2015, p. 318), cannot be underestimated when it comes to those students being able to reason about the data and thus work toward their own appreciation of foundation concepts including variation.

Finally, a natural progression in reasoning about variation is through reasoning about distribution (Reid & Reading, 2008; Shaughnessy, 2007), and clear learning links between the two concepts should be nurtured.

5.4 Reasoning About Distribution

This section discusses research on student reasoning about distributions of data and empirical sampling distributions. Historically, research on students’ general reasoning about data began with attention to separate statistical concepts such as average, variability, samples, and graphs. Reasoning about distributions requires a research design and tasks that allow for a more integrated investigation of student thinking about multiple statistical concepts such as shape, center, and spread and their relationships to one another. The word “distribution” is used to refer to a number of different entities in statistics and therefore can have multiple meanings. For example, there has been research into student reasoning about distributions of sample data, student reasoning about sampling distributions (empirical or theoretical), and student reasoning about probability distributions (both theoretical probability distributions and those generated by simulations). This section begins with some perspectives on the concept of distribution followed by an analysis of the evolution of research on students’ conceptions of distributions. Based on results of student thinking and responses to empirical tasks, some researchers have attempted to model the conceptual development and growth of student thinking about distributions.

A discussion and comparison of some of those models and some reflections on possible next steps for researchers concludes this section.

5.4.1 Perspectives on the Concept of Distribution

Perspectives in the literature view distribution as a meta-concept that is comprised of a number of statistical concepts about data, the most important of which is variation (spread, range, mean-absolute deviation,[2] standard deviation). Many of the other statistical concepts involved in distribution can be considered as general aspects of centers (mean, median, mode) or shape (clumped, symmetric, skewed, outlier, etc.). Reasoning about distributions involves integrating reasoning about these multiple statistical concepts.

Without variability in data, there is nothing to be distributed. In this regard, variability leads a learner to encounter the more encompassing concept of a distribution. According to Pfannkuch and Reading (2006, p. 4), "reasoning about distribution involves interpreting a complex structure that not only includes reasoning about features such as center, spread, density, skewness, shape, outliers but also involves other ideas such as sampling, population, causality and chance." After working with middle school students on tasks that involved comparing several data sets, Bakker and Gravemeijer (2004) claimed that distribution is the central concept for thinking about variability. In the introductory article of a special issue of the *Statistics Education Research Journal* on research on the concept of distribution, Wild noted that the concept of distribution "underlies virtually all statistical ways of reasoning about variation (Wild, 2006, p. 11)" and called distribution the lens used to view variation. Wild helps to clarify the concept of distribution by defining and discussing different types of distributions. Empirical distributions are frequency distributions where variability can be noticed directly in the data. Theoretical distributions are models that generate variation that is similar to what may be noticed in empirical distributions. Students are usually first introduced to reasoning about sample distributions of data that involve unit-to-unit variability, then to population distributions, and finally to sampling distributions that involve study-to-study variability (Wild, 2006).

The concept of distribution is so central to all of statistical thinking and reasoning that it is quite understandable how it has become a major focus of recent research in statistics education. Among the questions of interest to researchers are how and when do students begin to integrate the concepts of center, shape, and spread and to acknowledge that these are aspects of distributions? Is there a developmental sequence as students grow in their conceptions of distribution, and if so, what are some possible suggestions for teaching students and providing tasks about distributions that enable them to build on their conceptions over time? More recently the

---

[2] Mean absolute deviation, or MAD, is the (sum of the distances of all values from the mean)÷ (number of values). In analysis and measure theory, it is the L_1 norm.

issue of the teaching inference via simulations and sampling distributions has arisen and is being debated in the statistics education community. Following the development of teaching tools and recommendations about approaches to inference such as those found in Rossman and Chance (2014), there has been research into the efficacy of teaching inference using simulation-based approaches and sampling distributions (see, e.g., Lane, 2015; Taylor & Doehler, 2015).

5.4.2 Evolution of Research on Student Reasoning About Distribution

Research on student reasoning about distributions emerged from research on student reasoning about other statistical concepts such as centers, variability, samples, and graphs. (Research on student reasoning about centers and averages is discussed in Chap. 4 of this book; research on students' conceptions of variability is presented in the previous section of this chapter.) Previous research on centers and variability has helped to inform subsequent research on students' conceptions of distributions and has influenced the types of tasks used and types of research questions that have been explored around distributions. For example, a common theme that has emerged across research studies on student reasoning about average is that students' conceptions of average develop over a long time and that there appears to be a natural progression in student reasoning about averages from *mosts*, to *middles*, to the use of the mean as *typical* or as a *fair share*, to a *representative* for an entire data set (e.g., Konold & Pollatsek, 2002; Mokros & Russell, 1995; Watson & Moritz, 2000). Similarly, developmental hierarchies describing a progression for the variability concept were discussed above in Sect. 5.3 (Reading, 2004; Watson et al., 2007; Watson & Kelly, 2006), and a trajectory of conceptions of variability was summarized (Shaughnessy, 2007).

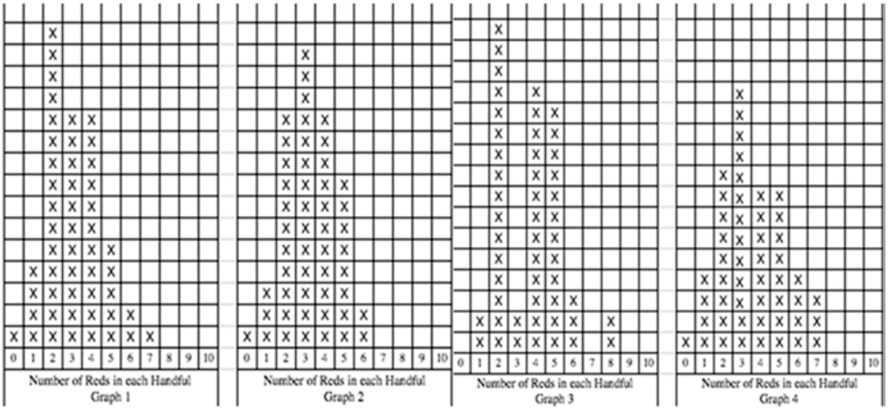
The research on student thinking on centers and variability revealed a tension for students between centers and variability when responding to tasks involving comparing data sets or sampling tasks. This tension was the first sign that there would be an added dimension of complexity for students as they attempt to integrate multiple statistical concepts in their development in reasoning about distributions. For example, when pulling repeated samples from a mixture of colored objects where they were given the population proportions of each color, some students tend to predict samples identical to or very close to the population proportions, while other students acknowledge there could be some, even considerable, variation in the sample proportions of the colors. Such tension between representativeness and variability has been documented by many researchers (Noll, 2011; Noll & Shaughnessy, 2012; Rubin, Bruce, & Tenney, 1991; Watson et al., 2007). Furthermore, a robust developmental hierarchy in students' integration of the concepts of expectation and variation has consistently been found (Saldanha & Thompson, 2003; Shaughnessy et al., 2004; Watson et al., 2007). Some students focus only on centers and expectations, while others focus only on variability in responses to sampling tasks. Still other students identify potential influences of both center and spread on samples, while a few students are not only able to recognize both expectation and variability but also to integrate the two concepts when reasoning about data that is gathered from repeated samples. Studies such as those by Saldanha and Thompson (2003), Reading and Reid (2006), Watson et al. (2007), Watson (2009), and Noll and Shaughnessy (2012) have provided evidence of what has come to be called *distributional reasoning* in the literature, the ability to identify and connect multiple aspects of distributions when reasoning about data (Shaughnessy, 2007).

Among the first researchers to take a distributional approach to student reasoning about data were Bakker and Gravemeijer (2004) who investigated seventh grade students' attention to center, spread, and shape when comparing distributions of data. Students tended to reason from particular data points and then move to *modal clumps* of data (Konold et al., 2002), referring to low clumps, average clumps, and high clumps in the graphs. Aspects of center, shape, and spread are evident in this type of student reasoning since the clumps (centers) are spread out across the distribution, creating a whole shape for the distribution of data. Friel, O'Connor, and Mamer (2006) also found that elementary students were likely to first focus on particular values such as the mode and to identify clumps of data when comparing data sets. Watson (2009) shared similar evidence of students referring to clumps. Konold et al. (2015) discussed four different perspectives on data: data as pointer, data as case value, data as classifier, and data as aggregate.

Subsequent researchers have used sampling tasks that attempt to provide opportunities for students to demonstrate that they are capable of distributional reasoning and to document how students respond to tasks that have opportunities to attend to multiple aspects of distributions, including shape, center, and variability. Some studies have asked students to draw repeated samples, either from a known or an unknown population. When students know the original population proportions, they are asked to use that information to predict what an empirical sampling distribution would look like for a statistic from the repeated samples (Shaughnessy et al., 2004; Watson & Kelly, 2006). In the case where the parent population is unknown, students are asked to use the information from the samples to predict the makeup of the original population. For example, Noll and Shaughnessy (2012) presented students with four empirical sampling distributions that had all been generated from the same binomial parent population (see Fig. 5.2). In this mystery mixture task, middle and secondary students ($N = 236$) were asked what they thought the original binomial proportions were in the parent population and how they made their decision.

The researchers wondered whether students would attend to variation both within and between the sampling distributions. Would they rely on visual modal clumps as centers or calculate medians or means? Would they incorporate both centers and spreads in their analysis to make estimates for the true proportion in the mixture? Results indicated that while students used a number of these possible strategies to make estimates for the mystery mixture, over 40% of them relied on "mosts," or modal clumps which tended to make them underestimate the true proportion of reds in the mixture (Noll & Shaughnessy, 2012). However, there were also many students

The graphs below all came from a class that is trying to estimate a mystery mixture of 1000 red and yellow candies in a large jar. They pulled 50 samples of size 10 (recording the number of reds and then replacing and remixing each time). Here are the graphs for the numbers of reds for four groups from that class.



a) What do you think the mixture in the jar might be?
b) Explain why you think this.

Fig. 5.2 The mystery mixture task

who articulated the variation evident among the sampling distributions and even some who calculated means to get information beyond the visual impact information in the graphs.

Bakker and Gravemeijer (2004) note that it is desirable for students to reason about aspects of distributions from both an upward perspective (from data to distribution) and from a downward perspective (from distribution to data, such as from a probability distribution to sample data). They recommend delaying the introduction of the mean with students, suggesting that a better approach is to build on the students' reasoning about modal clumps. Makar and Confrey (2005) claimed that "… there are more than just the two perspectives of distribution that are usually discussed in the literature: the single points and aggregate perspectives. This third perspective, … —partial distributions or "mini-aggregates"—deserves some further research to investigate the strength of its link to statistical thinking about distributions" (p. 48). The results of Noll and Shaughnessy (2012) found that although many students reasoned from such mini-aggregates, there were also more sophisticated responses that could be classified as reasoning from proportions or reasoning distributionally. Thus, the research on student reasoning about distributions has suggested that there may be a developmental-conceptual trajectory in student reasoning about distributions, similar to the trajectories that have been found when students reason about centers or variability.

5.4.3 Conceptual Models of Student Reasoning About Distribution

This section discusses four examples of models that describe trajectories of student reasoning about distributions. In part these models could be considered developmental, because as students mature their reasoning about data can include more abstract concepts such as measures of center and measures of variability that are not yet accessible to younger students. However, there is also a teaching-learning component to reasoning about distributions, because concepts such as mean, median, interquartile range, and standard deviation do not just emerge developmentally; they are taught. Thus, it may be more accurate to refer to these models as conceptual models that describe progressions of student reasoning that have emerged from research on student thinking about distributions and aspects of distributions (center, shape, and variability).

Ben-Zvi (2004) investigated students' reasoning about variability as they compared data sets. Ben-Zvi found that the students' progression of reasoning began with attention to variability, but progressed in the later stages to a consideration of multiple aspects of the data sets, including integrating measures of center and spread, and eventually to a concern about shape, and outliers in the data. Ben-Zvi noted that the "development of reasoning about variability in comparing the groups was accompanied by somewhat parallel development of global perception of a *distribution* as an entity that has typical characteristics such as shape, center, and spread" (p. 57).

Work to build a SOLO Taxonomy of tertiary students' reasoning about distribution was initiated by Reading and Reid (2006). Student tasks included comparing data sets, reasoning about samples pulled from a mixture of known proportion, a task on the behavior of the sampling distribution of means, and a task on comparing distributions. In their analysis Reading and Reid identified two U-M-R learning cycles. The first cycle focused on aspects of distribution (center, variability, shape), while the second cycle included attention to inference. In the first cycle, responses that made connections between several concepts (e.g., center and spread) were coded as relational (R1), indicating the beginning of distributional reasoning. In the second learning cycle, the researchers looked for connections between inferential statements and aspects of the distribution of data. Reading and Reid claimed that students' distributional reasoning depended heavily on the strength of their understanding of variation. They concluded that an understanding of variation may be a necessary condition for students to attain a deep understanding of distribution.

An investigation of Grade 3–9 students' reasoning about expectation and variation and students' integration of these two aspects of distributions was conducted by Watson et al. (2007). Student tasks included making predictions for repeated samples of lollies (candies) drawn from a known mixture, a two-spinner game, data from daily temperatures of the weather, and the comparison of pairs of distributions of student grades using both same-sized and different-sized samples. The study focused on links students made between expectation and variation and how the integration of the two concepts develops across grade levels. Watson et al. (2007) identified six levels of reasoning. The first two levels indicate either no or very primitive acknowledgment of expectation or variation. The top four levels suggest there is a developmental-conceptual progression in student reasoning about both expectation ("more" → centers in context → proportion → distribution with strong connections to variation) and variation (anything can happen → random variation → unexpected variation). Watson et al. (2007) identified a third developmental-conceptual progression of gradually stronger acknowledgment by students of the statistical links between expectation and variation. They recommended that tasks such as the ones they used in their study should be given to students since they provide opportunities for students to encounter and reason simultaneously about expectation and variation.

Noll and Shaughnessy (2012) researched Grade 6–12 students' understanding of empirical sampling distributions before and after a teaching episode in which students had opportunities to predict the results of repeated samples. Their tasks included repeated sampling from both known mixtures (prediction tasks) and unknown mixtures (the mystery mixture task) to estimate unknown population proportions. Students were asked to explain the reasoning for their predictions. Responses indicated that their reasoning was based on the shapes, centers, and/or variability that students expected in samples from the known distribution. Noll and Shaughnessy proposed a developmental-conceptual progression of student reasoning about sampling distributions that includes:

Level 1 *Additive reasoning*—using only frequencies to make predictions
Level 2 *Transitional reasoning*—attention to a single aspect of distribution such as shape, weak centers (modes or modal clumps), or spread
Level 3 *Proportional reasoning*—reasoning using means, medians, relative frequencies, or probability to make predictions
Level 4 *Distributional reasoning*—acknowledging and integrating multiple aspects of sampling distributions, shape, centers, and variability, when making predictions about sampling distributions (see Fig. 5.3)

All four of these models of conceptual development around reasoning about distributions identify similar conceptual trajectories that begin with reasoning from frequencies, particular data points, or modal clumps, then to relying on multiple aspects of distributions, such as center and spread, and eventually to integrating multiple aspects while reasoning about distributions. Reading and Reid (2006) call this final stage "relational thinking" in their SOLO model interpretation of student reasoning. Noll and Shaughnessy (2012) referred to it as distributional reasoning. Ben-Zvi (2004) called it "a global perception of a distribution" with characteristics such as centers, shape, and spread. Though there are some differences in the language used, researchers on student reasoning about distribution appear to be in general agreement that there is a progression of reasoning about distributions among school-age students similar to the progression from "mosts" to centers to proportions to distributions identified by Watson et al. (2007) and corroborated in the conceptual reasoning lattice of Noll and Shaughnessy (Fig. 5.3).

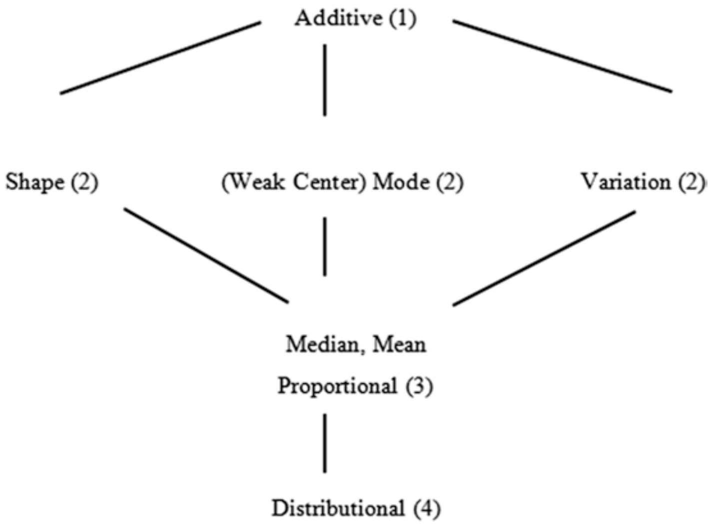


Fig. 5.3 Conceptual lattice of student reasoning about sampling distributions

In addition to the research studies of elementary, middle, and secondary students' reasoning about distributions and their aspects, there have also been studies on how tertiary students reason about distributions. The researchers used tasks similar to or even identical to those used with school-age students. Prospective elementary teachers' reasoning about variability and distributions was investigated by Canada (2006). Reasoning about distributions by introductory undergraduate statistics students has been investigated by Ciancetta (2007) and that of graduate mathematics students by Noll (2011). It is interesting to note that that the tendency found among upper secondary students to rely on proportions, centers, or probabilities to predict outcomes for repeated samples persists among tertiary students. Students tend to neglect variability in their analysis of and predictions about data. The studies by Canada (2006), Ciancetta (2007), and Noll (2011) provide further evidence for the conceptual trajectory of student reasoning about distributions that begins with students focusing on frequencies and then moving to making predictions based on proportions and relative frequencies and eventually to integrating various aspects of distributions, such as including both expectation and variability in their reasoning.

5.4.4 Reflections and Future Developments

Research on student reasoning about data started with studies of student reasoning about centers, moved to a focus on reasoning about variability in data, and eventually moved to investigations of student reasoning about distributions. Thus, the history of the research on student reasoning about data sets and sampling has come from looking at how students reason about particular aspects (center, spread) and then only later to global entities like distributions. Most of the research studies on student reasoning about distribution thus far have been exploratory studies, primarily descriptive in nature. However, the results of these descriptive studies have been consistent in their findings about how students reason about aspects of distributions (discussed in the previous subsection above). Models of student reasoning about distributions suggest that there may be transitions through which students must pass in their reasoning in order to fully comprehend the concept of a distribution as an entity. For example, proportional reasoning is a necessary condition prior to distributional reasoning. Students need to make predictions, interpretations, and inferences from distributions based on relative frequencies, rather than just from frequencies. Furthermore, a solid acknowledgment and understanding of variability around a center of a distribution is a critical piece of distributional reasoning.

The current state of research on distributions suggests several areas of needed research for the future. Based on the spadework provided by these descriptive studies, it may be an ideal time for statistics education researchers to begin to incorporate statistical tasks that target the hypothesized reasoning transitions about distributions into their teaching, to build teaching-learning trajectories for the classroom and test them. In this regard, the field is in need of curriculum design experiments (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) that can facilitate the development of student learning about distributions while simultaneously testing the validity of the proposed conceptual-developmental models. For example, it would be interesting to investigate the growth and changes in student reasoning about data that could occur if the teaching approaches recommended by Watson et al. (2007) and implemented for a short time by Noll and Shaughnessy (2012) were actually implemented throughout an entire introductory course in statistics.

Another area for future research is to conduct experimental studies that test novel teaching approaches to reasoning about data and distributions. The field may now have sufficient information from descriptive studies to conduct true experimental studies. In particular, the research on student reasoning about distributions has led to a growing interest among researchers in investigating students' informal inference, including inference based upon simulations of sampling distributions. Recent studies by Taylor and Doehler (2015) and Lane (2015) suggest the potential benefits of teaching inference using empirically generated sampling distributions. On the other hand, concerns have arisen about the advisability of incorporating inference from sampling distributions into introductory statistics courses (Watkins, Bargagliotti, & Franklin, 2014). It is high time that statistics education research conducted some experimental studies to test and compare various approaches to teaching inference in statistics classes. In particular, there is need for studies that compare the impact on students' statistical reasoning of informal inference approaches with that of traditional approaches to inference. The field is in need of both design experiments and true experimental studies to push beyond what is already known from the descriptive studies about students' reasoning about distributions. In any case, the growing interest in teaching reasoning from sampling distributions is likely to rekindle further research into student's understanding of distribution in general and hopefully will continue to build upon the prior research discussed in this section.

5.5 Reasoning About Comparing Groups

Previous sections of this chapter have elaborated on student reasoning about data (Sect. 5.2), student reasoning about variability and variation (Sect. 5.3), and student reasoning about distributions (Sect. 5.4). Comparing groups includes all of these aspects. Many statistical questions, hypotheses, and investigations in data are related to differences and similarities between groups. This may be one reason why Konold and Higgins (2003) see comparing groups as "the heart of statistics" (p. 206). Comparing distributions is fundamental in statistics since it includes many of the key ideas (Burrill & Biehler, 2011) like data, variation, and distribution, and it could therefore be considered as the most important activity in statistics education. Such activities can be done at different age levels ranging from learners at early stages (e.g., Watson & Moritz, 1999) to secondary school students (e.g., Pfannkuch, 2007) to pre- and in-service teachers (e.g., Makar & Confrey, 2002). Digital tools can play a fundamental role in group comparisons since they enable learners to easily change between different displays and summary statistics for getting a deeper insight into the data.

This section outlines a distinction between different facets of comparing groups and provides an outline of research about comparing groups across all age levels and about the use of digital tools when comparing groups. Finally this section offers some ideas for future developments.

5.5.1 Making Group Comparisons

When comparing groups at least two variables are taken into account. Three types of questions can be distinguished leading to three kinds of comparisons involving two variables (Konold, Pollatsek, Well, & Gagnon, 1997). The first kind of question,[3] for example, "Are males or females more likely to have a driver's license?" (Konold et al., 1997, p. 7), asks about the association between two categorical variables. A "group comparison" involves one numeric and one categorical variable, such as the question "Do those with a curfew tend to study more hours than those without a curfew?" (Konold et al., 1997, p. 7). The third kind of comparison involves two numeric variables, such as a question like "Is there a relation between hours spent watching TV and school grades?" (Konold et al., 1997, p. 7). This section considers the second kind of question, group comparisons. Reasoning about the association of

[3] For a more comprehensive categorization of statistical questions, see Biehler (2001, p. 98) and also Arnold (2013).

two categorical (first case of Konold et al., 1997) and two numerical variables (third case of Konold et al., 1997) are covered in Sect. 5.6 of this chapter.

The research tasks that have been used in studies that examine how students make group comparisons vary along several dimensions. One of these dimensions is the task setting. Four general types of task setting can be identified:

1. Whether or not the groups are of equal sample size
2. Whether or not the sample sizes are small or large
3. Whether or not the task requires the use of software to manipulate displays and calculate summaries to make comparisons
4. Whether the data set in which the group comparison takes place is a sample of a population or the population itself

Teachers and researchers have to be aware of these specific settings when designing tasks for group comparisons or when evaluating learners' outcomes when comparing groups, since each type of task setting may evoke and need different strategies for learners to compare the groups. A second dimension distinguishes between the following types of questions:

- Decision questions (e.g., "Which group is "better?")
- Descriptive and exploratory questions (e.g., "What differences and commonalities can be found?")
- Hypothesis-driven questions (e.g., "Do girls tend to read longer than boys?")

A third dimension includes the elements of distributions (called group comparison elements) that students can take into account when comparing groups. Features of distributions can be center, spread, skewness, peaks and clusters, and outliers (Rossman & Chance, 2001; Zieffler, Harring, & Long, 2011), which can also be seen as fundamental elements for describing a distribution. However, at least two additional elements can be compared across distributions. Based on observations of children when working on group comparisons, Biehler (2001, 2007a) suggested making a distinction between so-called proportion (p-based) and quantile (q-based) comparisons. Boxplots, for example, invite learners to compare distributions quartile-wise, such as comparing the medians or the lower quartiles. In p-based comparisons a specific cut point can be chosen (e.g., 10 h) and the proportion of cases that are equal to or larger than 10 h is compared in both groups (see Biehler, 2001). The use of these group comparison elements may point to different views on data when comparing groups in the sense of "local view" (e.g., view on local data points), "global view" (e.g., view on global characteristics of a distribution like center, etc.), and an intermediate step (e.g., view on clumps, bumps, hills, etc.) between both of them. Observations of children show they often pick out just an interval such as "between 2 and 3 h" and compare frequencies, which would be classified as a local view. P-based and q-based comparisons constitute a global view on the data and are valid comparisons from a statistical standpoint. In conclusion center, spread, shift, skewness, p-based comparisons, and q-based comparisons can be viewed as sustainable elements for comparing groups.

Taking into account the center and spread of distributions, boxplots are powerful displays and offer many advantages, especially when comparing more than two groups, but students also find them difficult to understand and interpret (see Bakker, Biehler, & Konold, 2005; Lem et al., 2014; Lem, Onghena, Verschaffel, & Van Dooren, 2013). There are other studies (e.g., Bakker & Gravemeijer, 2004; Konold et al., 2002) that have looked at students making group comparisons by identifying modal clumps in the distributions and comparing distributions via modal clumps. Modal clumps can be viewed as an adequate preliminary stage, in particular for young learners, to identify the location and spread of a distribution and to compare distributions by identifying a shift between two modal clumps.

A problem arises: a standardized language has not been developed for doing such comparisons in descriptive statistics, whereas in inferential statistics specific tests and models can be specified for comparisons such as that one distribution is just a linear shift of the other which means the distributions are identical except with regard to a measure of center. So, group comparisons can turn out to be a challenging activity for learners, especially when they are embedded in complex data sets with many variables, several displays, or several summary statistics. Imagine that a learner has produced boxplots, histograms, and numerical summary tables of two distributions and that (s)he may be able to further manipulate diagrams with a digital tool. This opens many new options for very complex comparisons across diagrams and groups. Learners may apply what Wild and Pfannkuch (1999) called transnumeration (an ability to change displays to find patterns in the data). In the case of a multivariate data set with many variables, a group comparison may also point to the relevance of additional variables that were not originally considered in the group comparison. For research on students dealing with group comparisons in the context of projects with multivariate data, see Biehler (2005).

5.5.2 Research on Group Comparisons with Students

Research on the group comparisons made by students has focused on one of three main goals:

- Categorizing the quality and level of students' comparisons
- Identifying pitfalls and difficulties students have with group comparisons
- Identifying intuitive strategies and which strategies are productive for future learning

Unfortunately, the research studies are only partly connected and related to each other so that one cannot speak of a cumulatively growing body of knowledge and theory. In our review, we do not include studies for group comparisons in the context of teaching and learning formal inference, but see Hogan, Zaboski, and Perry (2015) for an interesting study on this topic.

An early study on comparing groups was done by Watson and Moritz (1999), who used the interview protocol of Gal, Rothschild, and Wagner (1989) to observe

Australian Grade 4–8 students when comparing two distributions in different settings. Students were given two distributions of test scores for two school classes in the form of stacked dot plots. A total of four group comparison situations were given. The student's task was to decide "Which class is better?" (Watson & Moritz, 1999, p. 151).

A description of the four sets of two distributions provided for the four group comparison situations (see Fig. 5.4) follows:

- Part (a): Equal-sized samples; all scores of one distribution are larger than the scores of the other distribution.
- Part (b): Equal-sized samples; most scores of one distribution are larger than the scores of the other distribution.
- Part (c): Equal-sized samples; both distributions symmetric and with the same center; one distribution with larger spread than the other.
- Part (d): Unequal sample sizes; both distributions have the same range; the larger sample has a symmetric distribution; the distribution of the smaller sample has negative skewness, higher mean, and slightly larger standard deviation.

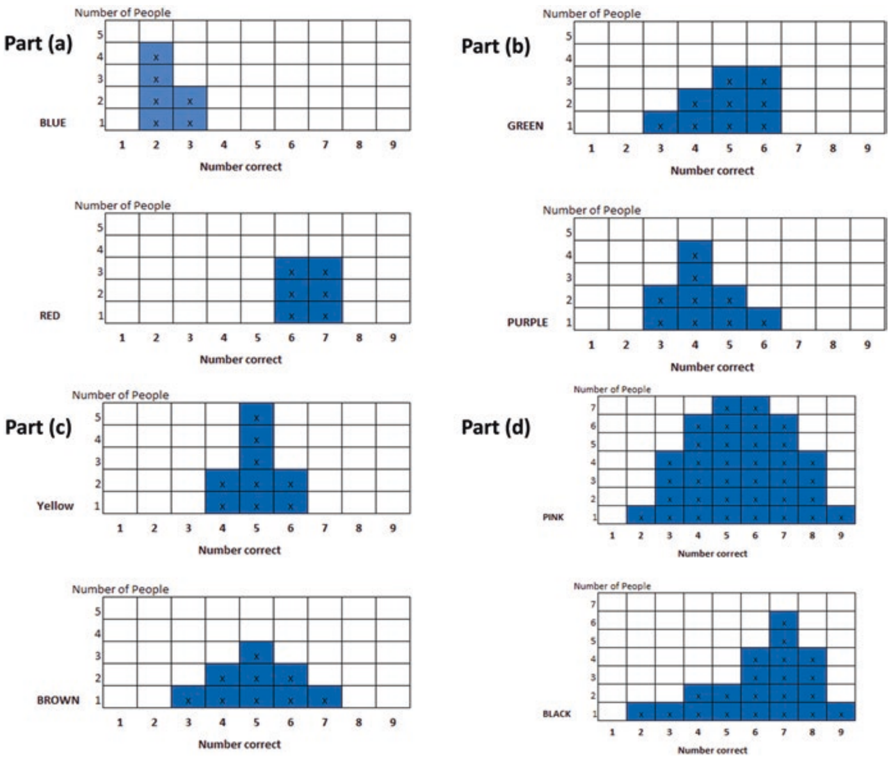


Fig. 5.4 Group comparison tasks similar to Watson and Moritz (1999, p. 151)—see also Watson and Shaughnessy (2004)

While a total score strategy (adding up all scores in both classes and comparing the sums) could be helpful to make decisions in Parts (a) and (b), it would not work for Parts (c) and (d) to decide which class is better. The responses of students were transcribed and coded based on the levels *unistructural*, *multistructural*, and *relational* of the SOLO Taxonomy of Biggs and Collis (1982). Strategies used by students when working on Parts (a)–(c) ranged from comparing individual values of the distributions to calculating the total of test scores, whereas calculating the mean of the scores of both classes was used when working on Part (d). One major result of the study was that students in higher grades tended to use strategies based on proportional reasoning compared to students in lower grades and that "students used numerical and visual strategies, either individually, or in conjunction with each other, to make comparisons between the data sets presented in graphs" (Watson & Moritz, 1999, p. 163).

While Watson and Moritz (1999) focused on young students without much preknowledge on comparing groups using numerical and visual strategies for comparing groups, Pfannkuch, Budgett, Parsonage, and Horring (2004) observed 15-year-old students when comparing distributions shown by boxplots with the temperatures of Wellington and Napier. According to Pfannkuch et al. (2004, p. 3), the

> students were required to pose a question (e.g., Which city has the higher maximum temperatures in summer?), analyze the data (given in the form of a table), draw a conclusion, justify the conclusion with three supporting statements and evaluate the statistical process. All students analyzed the data by calculating the five summary statistics with many using back-to-back stem-and-leaf plots for these calculations and then drawing boxplots by hand.

The researchers organized student responses into one of five categories: *conclusion*, *comparing equivalent summary statistics*,[4] *comparing nonequivalent summary statistics*, *comparing variability*, and *comparing distributions*. *Conclusion* responses were very general statements that made a group comparison, like "Napier has the highest temperature." *Comparing variability* responses took into account the comparison of variability of the two groups. Finally, *comparing distribution* responses included the comparison of distributions with regard to shift (see Biehler, 2004). The authors also assigned SOLO levels to the responses. There were no responses coded at the *relational* SOLO level and only a few at the *multistructural* level when comparing the boxplots via *comparing variability* or *comparing distributions*. This implies that the participants preferred to compare the boxplots via summary statistics (27 of 30 participants used summary statistics), but had difficulties using variability or shift.

Refining the framework of Pfannkuch et al. (2004), Pfannkuch (2007) asked Year 10 students to make three comparison statements to explain differences and similarities between the distributions of the variable "number of text messages sent in the last month" for customers from two phone companies (see Fig. 5.5).

[4] Summary statistics refer to measures of central tendency in this case.

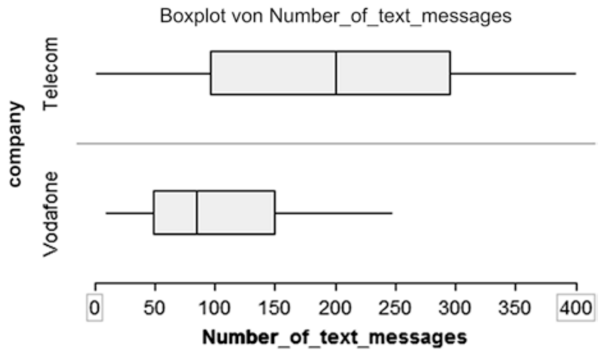


Fig. 5.5 Boxplot comparison task (copied) similar to Pfannkuch (2007, p. 157)

The categories of Pfannkuch et al. (2004) were refined in Pfannkuch (2007) to take into account students' responses on two different aspects. The first aspect identified concepts students used to compare distributions represented by boxplots, such as *summary* (e.g., comparison of medians), *spread* (e.g., comparison of range), *shift* (e.g., comparison of shift between both distributions), and *signal* (e.g., comparison of position and shift of the middle 50%), which can also be adapted for comparing distributions in other settings and with other displays. The second aspect rated the quality of the comparison on a four-point scale. A summary of the different level and group comparison elements given in Pfannkuch (2007) is reproduced in Table 5.1. A main conclusion was that the participants preferred to refer to summary and spread elements in contrast to shift and signal elements, and they tended to stay in the *describing* and *decoding* levels without interpreting their findings, seldom reaching the *assessor* level.

More recent studies on students' reasoning about comparing groups can be found in Langrall et al. (2011), Reaburn (2012), and Schnell and Büscher (2015).

For instance, the study of Langrall et al. (2011) focused on the role of students' context expertise when comparing distributions. Langrall et al. (2011) conducted a qualitative study and asked middle school students to analyze and compare authentic data, which was related to selected students' areas of interest (e.g., soccer, music, etc.). The authors found "that students used context knowledge to (a) bring new insight or additional information to the task, (b) explain the data, (c) provide justification or qualification for claims, (d) identify useful data for the task at hand, and (e) state facts that may enhance the picture of the data but are irrelevant to the process of analyzing the data" (Langrall et al., 2011, p. 47). Strategies of beginning university students when comparing two data sets have been investigated by Reaburn (2012). In this qualitative study, Reaburn (2012) asks the participants to state which group performed better (and why) in four different settings (Reaburn has used the tasks of Watson & Moritz, 1999, p. 151). Main results of the study were that the university students showed similar problems when comparing groups as younger students do (no use of measures of center, no proportional reasoning when appropriate). Schnell and Büscher (2015) analyzed individual concepts of students compar-

Table 5.1 Description of levels for student reasoning about comparing distributions by means of boxplots (Pfannkuch, 2007, p. 159)

Elements of reasoning	Point decoder[a]	Shape comparison describer	Shape comparison decoder	Shape comparison assessor
	Level 0	Level 1	Level 2	Level 3
Summary	Identifies the 5-number summary	Compares two or more corresponding 5-number summary points including median	Compares medians only. Compares non-corresponding 5-number summary points but does not interpret	Compares non-corresponding 5-number summary points and interprets
Spread		Compares spreads of visual shapes (lengths, spaces, in boxes) but does not decode	Compares and talks about spread, refers to range, compares local regions/densities	Compares and refers to the type of spread/ densities locally and globally (e.g., even, clustered)
Shift		Compares and refers to the shift qualitatively for the whole shape	Compares and refers to the type of shift (e.g., nonuniform)	Compares and refers to the shift of the majority
Signal		Compares the middle groups' visual shapes (boxes) in relation to each other but does not decode	Compares the middle groups and decodes by referring to the data	Compares the overlap of the data of middle groups

[a]The point decoder does not exist for spread, shift, and signal

ing distributions. The major goal of their research was to explore how German middle school students (ages 13–15) without specific preknowledge made comparisons of frequency distributions represented as stacked dot plots. The students' task was to compare the distributions of temperature readings taken at the top of the highest German mountain, Zugspitze, during July in 2002, 2004, and 2007. The individual concepts when comparing groups were reconstructed with the help of using an adaption of the framework of Vergnaud's (1996) theory of conceptual fields. This framework was applied to identify several concepts that students used for comparing the distributions. Schnell and Büscher (2015) observed that students used visual features, such as modal clumps, to organize the data and that the students primarily focused on absolute frequencies instead of the relationship of corresponding features between the distributions.

5.5.3 Research on Group Comparisons with Digital Tools

The focus in Sect. 5.5.2 has been on group comparison strategies of learners where given displays of distributions like stacked dot plots were provided (e.g., Watson & Moritz, 1999) or boxplots (e.g., Pfannkuch, 2007). We now focus on research on the use of digital tools when comparing groups (with preservice teachers and in-service teachers). Makar and Confrey (2002, 2004) and Makar (2004) reported observations from a professional development course where preservice teachers compared two groups using Fathom (Finzer, 2007). The preservice teachers were given dot plots of test scores from two different schools. In contrast to studies mentioned in the previous section, this task contained a sample/population setting where preservice teachers were given the data set in Fathom and were required to manipulate the data displays to decide whether the groups are different (Makar & Confrey, 2004). More precisely they were "asked to use Fathom to compare the performance of males and females in the school" (Makar & Confrey, 2004, p. 360). A five-tier framework was developed to categorize the teachers' reasoning when comparing two groups: *pre-descriptive*, *descriptive*, *emerging distributional*, *transitional*, and *emerging statistical* level. Reasoning at the *pre-descriptive* level was based on individual data points or anecdotal evidence. *Emerging distributional* reasoning consisted of a holistic view that used qualitative descriptors and summary statistics to describe two data sets. At the most developed level of reasoning, *emerging statistical*, the teachers began to consider the differences between measures of center in light of variability and to take sample sizes into account. In summary, the taxonomy of Makar and Confrey (2002) primarily had a focus on possible inferences learners can make when comparing groups with regard to variability between both data sets and statistical terms like "evidence" and "significance." Madden (2008) used the Makar and Confrey (2002) framework to examine preservice teachers' reasoning when comparing groups to show an improvement of teachers' skills in comparing groups after a professional development course. In a follow-up study of Makar and Confrey (2002), Makar and Confrey (2004) reported on a video interview of four pairs of preservice teachers after their participation in a professional development course in statistics. The preservice teachers were asked to use Fathom to work with a data set of test scores to decide whether the distributions were different for boys and girls. Makar and Confrey (2004) found that the teachers were comfortable using traditional descriptive statistical measures to conduct informal comparisons, although the teachers struggled to interpret the difference between variation within distributions and variation between groups.

We are now going to examine studies that focus closely on the influence of software on comparing groups and how preservice teachers use software when comparing groups. Based on studies with school children, Biehler (1997) conceptualized four phases when learners solve statistical problems with software: formulating the context problem as a *statistical problem*, transforming the *statistical problem as a task for the software*, *using the software*, and *interpretation of software results in terms of statistics*. The research report pointed out that quite a few students directly jump from a real problem to a task for the software without a careful consideration whether the problem changes during this process. Moreover quite a few students are satisfied with producing computer results that are neither interpreted in statistical nor subject matter terms (Biehler, 1997). One may conclude, when digital tools are used, comparing groups becomes an even more challenging activity, since there are many ways to use digital tools in group comparison processes. In this respect, Makar and Confrey (2014) established a typology for learners when doing a data analysis task with digital tools, in this case Fathom. The authors distinguished between three different approaches: *wondering*, *wandering*, and *unwavering*. According to Makar and Confrey, *wonderers* look on the data with a certain theory in their mind and can be characterized as goal-oriented, seeking "evidence to support, refine, and extend their theories" (p. 356). In contrast, *wanderers* "have no particular evidence in mind when going into the data" (p. 357). *Wanderers* explore the data to see if anything "popped out" at them. *Unwaverers* can be identified "by the decision pathway used: investigators looked for a particular piece of evidence to support or refute their original conjecture, and once they found it they were satisfied that they had answered the question put to them." (p. 357).

TinkerPlots (Konold & Miller, 2011) can also be seen as an adequate tool not only for primary but also for secondary students and for teacher education to enhance the quality of group comparisons, switching easily between several displays. In this regard Frischemeier (2017) designed, conducted, and evaluated a one-semester-long course for preservice teachers on data analysis with TinkerPlots via a design-based research approach (Cobb et al., 2003). One major aim of the course was to teach preservice teachers how to compare groups using TinkerPlots. Building on the findings of Biehler (1997), Frischemeier conducted a video study with the participants after the course (Frischemeier, 2014, 2017; Frischemeier & Biehler, 2016). The preservice teachers were given real data about the monthly income structure of German employees and were asked to compare the distributions of income of men and women in Germany using TinkerPlots. This was called the "VSE task" (see Fig. 5.6).

Communication and activities with TinkerPlots were analyzed. On this basis, Frischemeier (2014) identified and verified findings that Biehler (1997) reported for students working on group comparisons with TinkerPlots. The teachers' transcribed communications in their group comparison process were analyzed (Frischemeier, 2017; Frischemeier & Biehler, 2016) using a structuring and scaling qualitative content analysis approach (Mayring, 2015). In regard to the work of Pfannkuch (2007) and Biehler (2001, 2007a, 2007b), the preservice teachers were found to use many of

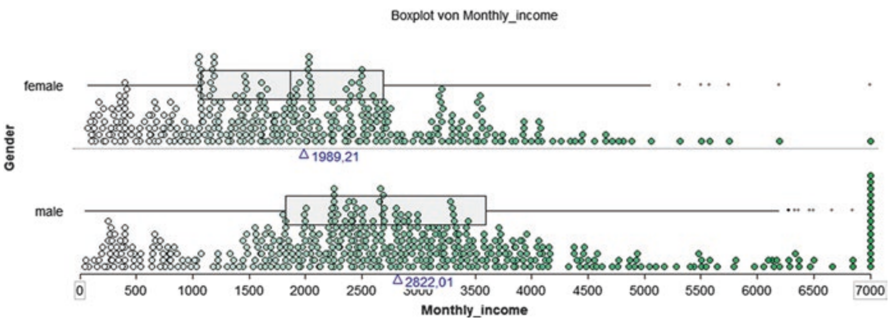


Fig. 5.6 VSE task—boxplots of the distributions of the variable "Monthly_income" separated by "Gender" (Frischemeier, 2017, p. 515)

the elements identified earlier in Sect. 5.5.1. In the study the level of each teachers' reasoning was rated as high, medium, or low. A medium reasoning level was given when a difference was described (typically, "the mean of male employees is higher than the mean of female employees"). Reasoning that went a step further and interpreted the differences (typically, "male employees earn more than female employees on average") was coded as high. A low coding was given when differences were worked out incorrectly. Major results of the study were that the participants used a broad spectrum of comparison elements, but overall showed at least only medium reasoning levels (no interpretation) on the elements they used for comparing groups.

One major finding with regard to the TinkerPlots use and the TinkerPlots skills of the preservice teachers when working on the VSE task (see Fig. 5.6) was that the participants showed high software skills and could conduct their statistical investigations with TinkerPlots, but tended to neglect the interpretation of their graphs produced with TinkerPlots. Implications and recommendations for learning trajectories for group comparisons for preservice teachers with TinkerPlots are to focus on the interpretation of findings by discussing adequate and non-adequate examples with peers and in the classroom. Furthermore, a data analysis scheme may help learners to structure their findings in a complex data analysis exploration (Frischemeier, 2017; Frischemeier & Biehler, 2016).

Before introducing students into the use and the comparison of boxplots, hat plots, which divide the data set in three areas (lower 25%, middle 50%, upper 25%), can offer an adequate preconcept for learners. Details on the use of hat plots can be found in Watson, Fitzallen, Wilson, and Creed (2008).

5.5.4 Reflections and Future Developments

This section distinguished several settings when comparing groups, identified some approaches in evaluating learners' processes when comparing groups, and described the use of different frameworks: SOLO (Pfannkuch et al., 2004; Watson & Moritz, 1999), five-tier framework (Makar & Confrey, 2002), and structuring and scaling approach (Frischemeier & Biehler, 2016; Pfannkuch, 2007). There are several points that might be interesting for further research. First, research on how learners compare groups in the contexts of large open multivariate data sets are needed, especially with respect to how learners use software to determine differences between two distributions. Second, there is a need for the development of teaching and learning material to support the development of students' skills in the interpretation of findings in a data analysis (or group comparison process). Third, there is a need to develop supporting material like data analysis schemes, which refer to sustainable group comparison elements like center, spread, skewness, shift, and p-based or q-based comparisons, that might be helpful to structure learners' processes when comparing groups and to support tool use and exploration process. Another aspect worthy of further research related to the third aspect would be research on the documentation of findings of the data exploration process—especially when exploring multivariate data with digital tools. Here the extraneous cognitive load of learners is high, because they have to concentrate on multivariate data, software use, and interpretation and documentation of findings.

5.6 Reasoning About Covariation and Associations Between Variables

In Sect. 5.5 three types of association (see Konold et al., 1997) were distinguished: categorical vs. categorical, numerical vs. categorical, and numerical vs. numerical. Whereas so-called group comparisons are numerical vs. categorical as outlined in Sect. 5.5, the focus in this section is on the categorical vs. categorical and numerical vs. numerical types. In summary, this section considers research on learners' reasoning about covariation and association in different settings (e.g., contingency tables, scatterplots, scatterplots with superimposed lines or curves), as well as examples of software tools that might enhance learners' reasoning about association and covariation.

The section first looks at research that focuses on understanding contingency tables and scatterplots. Second, different frameworks for evaluating learners' reasoning about covariation are considered. Finally, reports on the use of specific digital tools to enhance statistical reasoning when taking into account association and covariation are outlined.

5.6.1 Reasoning About Association and Reasoning About Covariation

Reasoning about association involves "the analysis of contingency tables, the determination of correlation between quantitative variables, and the comparison of a numerical variable in two or more samples" (Batanero, Estepa, Godino, & Green, 1996, p. 151).

The first two types of association are the focus of this section as the third type was discussed previously under group comparisons. One important remark is that although one may want to find causal explanations that allow understanding of the environment, association does not necessarily indicate a causal relationship (Batanero et al., 1996). Further definitions of association can be found in Carlson, Jacobs, Coe, Larsen, and Hsu (2002), Zieffler and Garfield (2009), and Moritz (2004).

As in other fields of reasoning about data, data displays and tables are crucial to enhance learners' insight into association. A paradigm display for type (1) association is contingency tables and for type (2) scatterplots, but neither of these kinds of displays are straightforward for students. Whereas different percentages (e.g., cell, column, and row) in tables can be difficult for learners to interpret, scatterplots "provide information about two variables that are not necessarily dependent on each other and show the correspondence of the ordination of each variable" (Moritz, 2004, p.40 cited in Fitzallen, 2012). Difficulties for learners in decoding contingency tables or scatterplots are also identified by Konold (2002). A basic problem that amplifies the difficulties is that understanding association requires the coordination of deterministic conceptions (functional relationships) with statistical variation. This can be an obstacle for learners as pointed out, for example, by Noss, Hoyles, and Pozzi (2002).

As an overview, Garfield and Ben-Zvi (2008, p. 299) outlined the following general findings on developing covariational reasoning and some prevalent conceptions and misconceptions of learners:

- "Students' prior beliefs about the relationship between two variables have a great deal of influence on their judgments of the covariation between those variables;
- Students often believe there is a correlation between two uncorrelated events (illusory correlation);
- Students' covariational judgments seem to be most influenced by the joint presence of variables and least influenced by the joint absence of variables;
- Students have difficulty reasoning about covariation when the relationship is negative;
- Students' covariational judgment of the relationship between two variables tends to be less than optimum (i.e., smaller than the actual correlation presented in the data or graph);
- Students have a tendency to form causal relationships based on a covariational analysis."

5.6.2 Reasoning About Association in Two-Way Contingency Tables

There is plenty of research on the reasoning about association in contingency tables beginning with Inhelder and Piaget (1955) and continuing with Batanero and colleagues to recent research from Watson and colleagues. Batanero et al. (1996) and further studies (Batanero, Estepa, & Godino, 1997; Batanero, Godino, & Estepa, 1998) have built on the research of Inhelder and Piaget by observing students' reasoning about associations in contingency tables. Specifically, they conducted research with 213 pre-university students (without specific prior knowledge of the domain) who were given a questionnaire with five tasks in which the students were asked to identify associations among given variables. The five tasks (see Batanero et al., 1996) varied in certain aspects: type of table (2×2, 2×3, 3×3), sign of the association (direct, inverse, independence), and relationship between context and prior belief (prior belief ("theory") agrees with data, prior belief goes against data). The example in Fig. 5.7, taken from Batanero et al. (1996), is classified as "independent" and "theory contradicted."

Item 1 (Smoking). In a medical center 250 people have been observed in order to determine whether the habit of smoking has some relationship with broncial disease. The following results have been obtained.

	Bronchial disease	Non bronchial disease	Total
Smoke	90	60	150
Not smoke	60	40	100
Total	150	100	250

Using the information contained in this table, would you think that, for this sample of people, bronchial disease depends on smoking? Explain your answer.

Fig. 5.7 Display similar to Batanero et al. (1996, p. 168)

Responses to the questionnaire items were coded for the type of association perceived by the students (direct association, inverse association, or independence). Students' strategies were categorized using the framework of Pérez Echevarría (1990), which categorizes students' strategies for 2 × 2 contingency tables (see Fig. 5.8 for format) into one of the following five levels:

- Level 1: Using only one cell in the table, usually cell [a]
- Level 2: Comparing [a] with [b] or [a] with [c]
- Level 3: Comparing [a] with [b] and [a] with [c]
- Level 4: Using all four cells in the table, employing additive comparisons
- Level 5: Using all four cells in the table, employing multiplicative comparisons (five levels of Pérez Echevarría, 1990, cited by Batanero et al., 1996, p. 154)

A major result of Batanero et al. (1996) was that the students showed good intuitive abilities for judging association in 2 × 2 contingency tables. A further result was that a large number of students were incapable of providing an argument for the 3 × 3 contingency table (33 cases out of 213) or were unable to make a judgment of association in this table (49 additional cases). Here it can be said that the task is more difficult when the dimensions of the table are increased.

There are also several incorrect strategies that were observed by Batanero et al. (1996) when students have worked on the task. The authors distinguish the following three types:

- Students with a determinist conception of association did not note exceptions to the existence of a relationship between variables and expected each value in the dependent variable to correspond to each value of the independent variable.

Fig. 5.8 Typical format for a 2 × 2 contingency table (similar to Batanero et al., 1996, p. 153)

	B	Not B	Total
A	a	b	a+b
Not A	c	d	c+d
Total	a+c	b+d	a+b+c+d

- Students with a unidirectional conception of association recognized dependence only for direct associations and considered an inverse association as representing independence.
- Students with a "localist" conception of association used only part of the data in the contingency table to make a judgment of association (Batanero et al., 1996).

In a subsequent research project, Batanero et al. (1997) described a teaching experiment that investigated preservice teachers' understanding of association. A test was administered before and after the preservice teachers completed a course that included numerous topics on covariation and association. One result was that most preservice teachers overcame deterministic conceptions and accepted randomness. Another conclusion from Batanero et al. (1996, 1997) is worth noting: while the preservice teachers moved from an additive approach to using multiplicative comparisons, they still lacked proportional reasoning (Batanero et al., 1996). A significant percentage of the students demonstrated good intuitive ability for judging association in 2 × 2 contingency tables, but had greater difficulty when the dimensions of the table were increased. Students demonstrated the causal misconception throughout the course (Batanero et al., 1996, 1997). Additionally, problems with distinguishing the role of independent and dependent variables and problems when reasoning about relationships that were negative were strongly evident in these studies (see also Zieffler & Garfield, 2009).

The ideas and problems of Batanero et al. (1996, 1997) have been used in more recent research. Watson, Callingham, and Donne (2008) administered an association problem (the lung disease problem) from Batanero et al. (1996, 1997) to students and used the students' responses to evaluate teachers' pedagogical content knowledge (PCK). The teachers were provided with student answers to the problem and asked which typical responses their students would give on the item, how the teachers would use the item in the classroom, and to write a response to the student answers. Similarly, Watson and Callingham (2014) administered the same association problem to 110 students from Year 6 to Year 11. Teachers were shown the association problem and asked in an interview to name the big statistical ideas that were embedded in the lung disease problem, give examples of appropriate and inappropriate responses of students, and state opportunities for teaching the given problem. Then the teachers were shown student responses and were asked to explain "how to move this student's understanding forward" (Watson & Callingham, 2014, p. 260). All hierarchies and rubrics for evaluating student and teacher reasoning and ensuing analysis are described in detail in Watson and Callingham (2014). Further research on school students reasoning when working on the lung disease problem and the indigestion problem can be read in Watson and Callingham (2015). Studies on preservice teachers' understanding of aspects of probability of basic events have been conducted by Estrada Roca and Batanero (2006) and Contreras, Batanero, Diaz, and Fernandes (2011). Research on primary school students' strategies in solving contingency table problems with special focus on the role of intuition and inhibition can be found in Obersteiner et al. (2015).

Examples of research on student and adult reasoning while engaging with Bayesian probability problems, which is also related to reasoning about associations in *m*x*n* contingency tables, include studies by Gigerenzer (1996, 2002), Wassner (2004), and McKenzie and Mikkelsen (2007). One major idea coming out of these studies is to use natural frequencies rather than relative frequencies when introducing Bayesian probability problems. Readers are referred to these studies for further detail of students' reasoning about Bayesian probabilities.

5.6.3 Reasoning About Associations Between Numerical Variables

This subsection refers to the third type (numerical vs. numerical) of association and presents two exemplary research studies, the study of Moritz (2004) focusing on reasoning about covariation among elementary and middle school students and the study of Zieffler and Garfield (2009) dealing with the reasoning about covariation of undergraduate students. In the study of Moritz (2004) with elementary and middle school students, three skills needed for reasoning about covariation among elementary and middle school students were named: speculative data generation, verbal graph interpretation, and numerical graph interpretation. Survey responses were collected from 167 students in Grades 3, 5, 7, and 9. Task 1 assessed speculative data generation by asking students to draw a graph for a verbally given covariation. Task 2 assessed both verbal and numerical graph interpretation by having students respond to questions related to a given scatterplot. The analysis categorized students' success in verbally generalizing the required covariation and numerically interpreting covariation into four levels: *nonstatistical*, *single aspect*, *inadequate covariation*, and *appropriate covariation*. According to Moritz (2004), numerical graph interpretation was highly correlated with both verbal graph interpretation and speculative data generation, whereas there was a weaker correlation between verbal graph interpretation and speculative data generation. Moritz also found evidence that all students were engaged in the task as most students could identify at least one aspect related to the data. Overall, Moritz identified two misconceptions/erroneous approaches and one difficulty learners face when reasoning about covariation: focusing on isolated bivariate points; focusing on a single variable in a bivariate data plot; and reasoning about a negative relationship that contradicted prior belief.

In their studies with undergraduate students, Zieffler and Garfield (2009) examined the development of reasoning about quantitative bivariate data in a one-semester introductory statistics course. One of their aims was to identify patterns of change in students' reasoning about quantitative bivariate data throughout the course. Three instruments were used to collect student data on four occasions during the course across four cohorts of students. Analysis of the data revealed that marked growth in reasoning about bivariate data happened primarily during the first unit of the course, before bivariate data was formally taught, and that students' reasoning about bivariate data did not increase in a constant linear fashion. They suggested that the development of students' reasoning about bivariate data may be generally related to their development of statistical reasoning rather than a result of formal instruction and is directly related to their reasoning about distributions.

Going one step further to the reasoning about correlation and regression, Engel and Sedlmeier (2011) identified three issues related to understanding correlation and regression: *psychological biases*, *mathematical difficulties*, and *difficulties with the functional understanding of associations*. *Psychological biases* occur when students have difficulties making judgments about associations for psychological reasons, such as, influence of previous beliefs. *Mathematical difficulties* are misconceptions based on mathematical content, such as the misconception that a high correlation does not imply validity of a linear model. *Difficulties with the functional understanding of associations* are unidirectional conceptions of association where learners consider only positive relationships to represent an association and inverse relationships to represent independence. The misconceptions across all three issues are listed in Table 5.2.

Many of the identified psychological biases and mathematical difficulties are related to a deterministic world view (Engel & Sedlmeier, 2011). To achieve a better

Table 5.2 Several misconceptions of students when thinking of correlation and regression (table similar to Engel & Sedlmeier, 2011, p. 248)

Psychological biases	• Influence of previous beliefs • Illusory correlation • Misjudgment of strength of covariation • Confounding variables • Regression effect regarded as real effect • Transitivity misconception
Mathematical difficulties	• Association instead of dependence • High correlation does not imply validity of a linear model • Interpreting the correlation coefficient
Difficulties with the functional understanding of associations	• Deterministic conception of association • Unidirectional conception of association • Local conception of association • Causal conception of association

appreciation of statistical associations and the role of variation of statistical data, they introduced different versions of the "signal and noise" representation of data (Fig. 5.9). These different interpretations of "signal and noise" are expected to counteract overreliance on deterministic thinking. Engel and Sedlmeier (2011) suggested that curricula need to be based on knowledge of students' potential fallacies and misunderstandings, should make use of real data and technology, and provide experience with modeling to overcome the shortcomings listed in Table 5.2.

Additionally, Engel and Sedlmeier (2011) offered several recommendations to overcome the shortcomings mentioned in Table 5.2. Educators should make use of fallacies and misunderstandings: real data, experience in modeling, and technology. Like other researchers, they saw the huge potential of the use of technology to overcome shortcomings when thinking about associations of data. As an example, they referred to residual plots that can easily be displayed in Fathom and that allow the user to investigate the deviation between model and data. Several software packages and tools have the potential to enhance the understanding of covariation and are discussed in the Sect. 5.6.4.

Regression models aim at modeling relationships between numerical variables. One special case of regression that can be taught in secondary school is linear regression. In regard to the association between students' conceptualizations of slope and students' understanding of the line of best fit, Casey and Nagle (2016) investigated in which way students are able to accurately place a line of best fit in given scatterplots relative to the least-squares regression line. They distinguished different slope conceptualizations of learners like *linear constant*, *behavior indicator*, *real-world situation*, *functional property*, *trigonometric conception*, and *physical property*. Definitions and examples of each conceptualization can be found in Casey and Nagle (2016). In a follow-up qualitative study, the authors conducted task-based interviews with seven students (Grade 8) concerning the placement and concerning their reasoning about the line of best fit in four different task settings, where scatterplots were given. One result of the study was that the conceptualization of slope of these students plays a significant role when reasoning about the placement of the line of best fit.

Statistics education has to develop concepts and supporting materials to help learners to deal with challenging topics like regression. Also identifying preliminary stages to difficult concepts or displays could be helpful. A good example of how to support the reading and interpretation of a scatterplot is given by Noss et al. (2002).

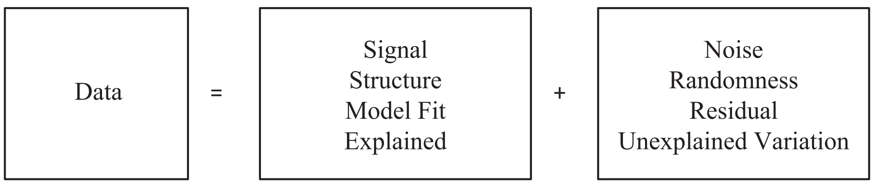


Fig. 5.9 Different versions of the signal-noise representation of data (display similar to Engel & Sedlmeier, 2011, p. 253)

The authors provided approaches to helping nurses gain a deeper understanding of covariation in scatterplots. Their approach is related to Konold's (2002) "sliced scatterplots" (see Fig. 5.11) method: slicing a scatterplot bridges the gap between statistical variation and deterministic dependence. In sliced scatterplots a continuous variable is reorganized into categories (see Figs. 5.10 and 5.11 in the next section). The intention of a sliced scatterplot display is that students can see "each vertical slice of data in this plot as a distribution of a discrete group, [and that] students can apply skills they have learned in comparing two distributions to visually compare the centers of the distributions in the sliced scatterplot" (Konold, 2002, p. 3). Technology is very helpful for changing scatterplots into sliced scatterplots and back again. Further possibilities of enhancing students' reasoning about association or covariation with digital tools follow in the next section.

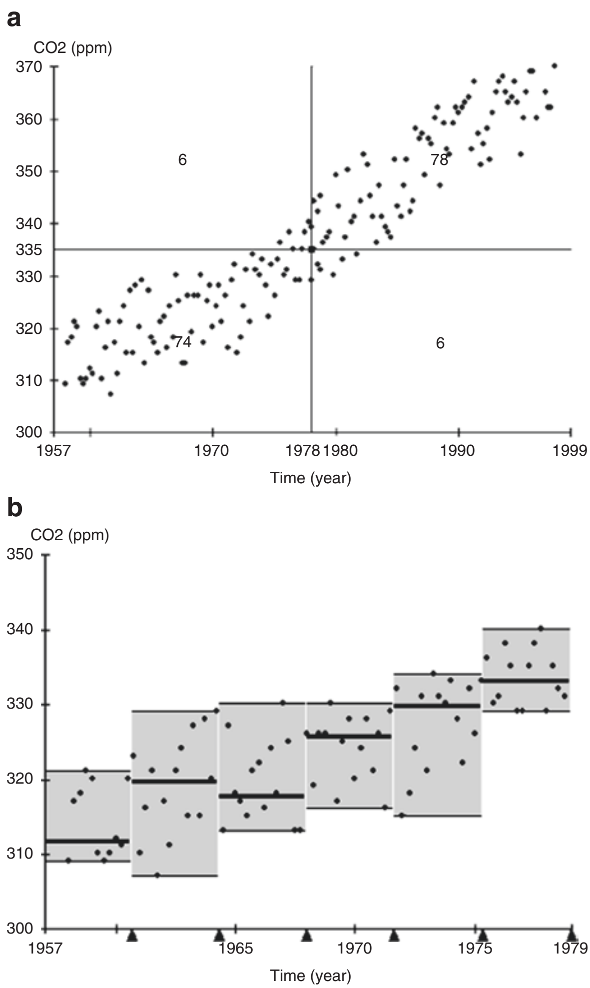


Fig. 5.10 Cross option (*left*, **a**) and the two equal groups option (*right*, **b**) of a minitool; figures similar to Cobb, McClain, and Gravemeijer (2003, pp. 18–19)

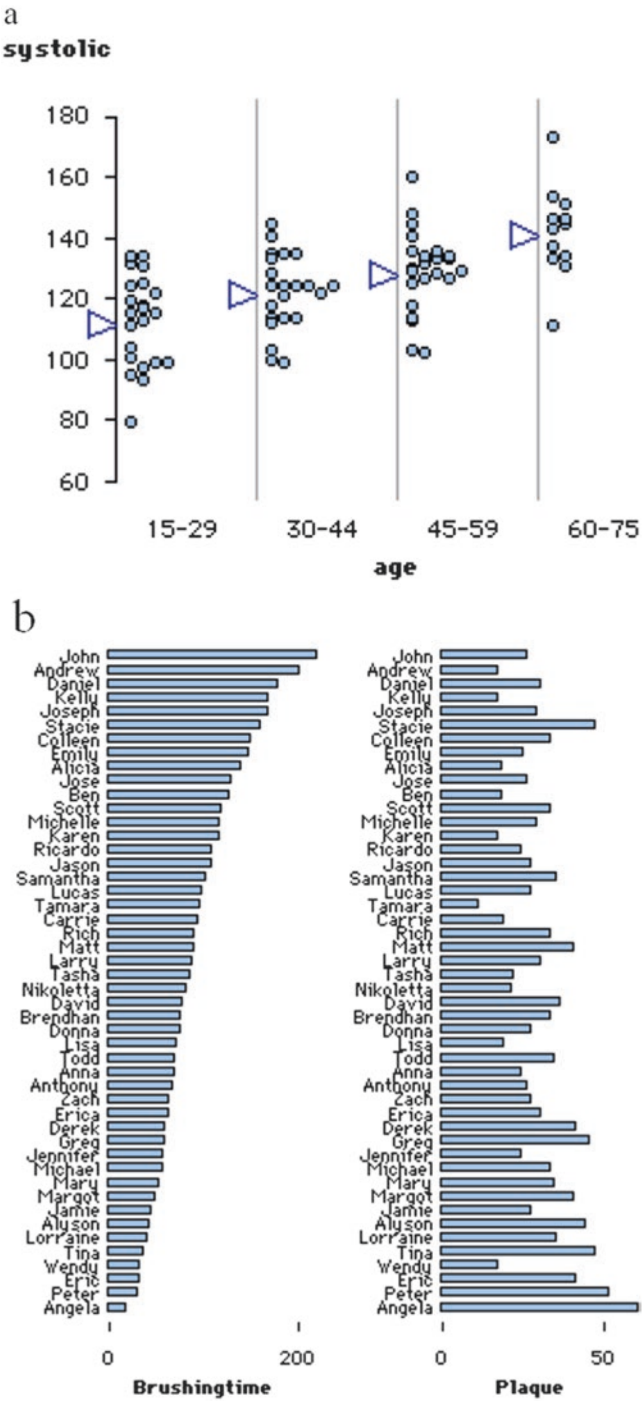


Fig. 5.11 Scatterplot slices (*left*, **a**) and ordered case value bars (*right*, **b**) similar to Konold (2002, pp. 2–3)

5.6.4 Technology to Enhance a Better Understanding of Association and Covariation

An overview of technology that supports learner's reasoning when analyzing data is given in Biehler, Ben-Zvi, Bakker, and Makar (2013). TinkerPlots (Konold & Miller, 2011) and Fathom (Finzer, 2007) support the exploration of association between variables in two-way tables by enabling the students to easily switch between column, row, and cell percentages. Similarly, Pfannkuch and Budgett (2017) introduced eikosograms to help students visualize association between variables and to build conditional probability concepts, respectively, to better visualize data from two-way contingency tables.

The use of technology may also enhance covariational thinking. Therefore it is important to know which digital tools are at hand and how these digital tools may enhance covariational thinking. For example, tools can help learners to create sliced scatterplots which were described in Sect. 5.6.3. Cobb, McClain, and Gravemeijer (2003) observed the covariational reasoning of students using minitools[5] for identifying patterns in the relationship between two numeric variables. The instruction guided students to the question of how the dependent variable changes "on average" with the independent variable. As an example, Cobb et al. (2003) used the investigation of association between time (year) and CO_2 (ppm). The minitools offered the possibility of displaying the data with the scatterplot divided into four equal-sized parts (see Fig. 5.10a). Another feature of minitools was the separation of several slices of the scatterplot into equal interval subgroups based on the independent variable (see Fig. 5.10b). Comparison of medians in each consecutive subgroup helps students see how CO_2 levels rise over time. Based on the results of the interview study, Cobb et al. (2003) suggested starting with the shape of distributions rather than with variability of distributions.

The valuable features of the minitools were deliberately integrated into the software TinkerPlots (Konold & Miller, 2011). It offers additional possibilities for gaining insight into covariation between two numerical variables. Like Cobb's minitools, TinkerPlots offers the option of displaying sliced scatterplots—in this case displaying the means (triangles in Fig. 5.11a) of the subgroup distributions. Using TinkerPlots, Konold (2002) explored the data of Noss, Pozzi, and Hoyles (1999) with sliced scatterplots (see Fig. 5.11a) to show how TinkerPlots can be used to support the conceptual development that Noss et al. were aiming for. Another example given by Konold (2002) is the exploration of association between brushing time and plaque when brushing teeth using ordered case value bars in TinkerPlots (see Fig. 5.11b).

Moreover, TinkerPlots offers a color gradient feature, which might be especially useful for younger students to represent and study an association between two vari-

[5] The development and the implementation of these minitools were the starting point to develop the educational software TinkerPlots, which includes the features of the minitools.

ables. For example, Konold (2002) investigated the relationship between the body length and age of cats. The TinkerPlots display in Fig. 5.12a does not imply a relationship between body length and age of cats (intensity of gray shade displays the age in years), since older cats can be found in the middle and on the almost right end of the distribution. However, when taking into account the variable gender, it can be seen (Fig. 5.12b) that male cats (white points) tend to have a larger body length than female cats (gray points).

The potential of TinkerPlots to enhance covariational reasoning was investigated in the PhD thesis of Fitzallen (2012). Fitzallen investigated learning sequences about covariation with TinkerPlots and the related learning processes of students. In student interview task settings, where students (Years 4, 5, and 6) had been working with TinkerPlots on exploring covariation in real data, three different strategies of these students could be observed (Fitzallen, 2012): *snatch and grab*, *proceed and falter*, and *explore and complete*. Whereas students following the *snatch and grab* strategy often neglect to interpret the data, students following an *explore and complete* strategy show the most sophisticated covariational reasoning with TinkerPlots because they used TinkerPlots effectively to interpret their graphs and successfully combined their knowledge on how to use TinkerPlots and how to interpret the graphs they generated with TinkerPlots.

Fitzallen (2012) conducted further research to investigate how students develop an understanding of covariation with TinkerPlots. Reasoning about covariation was rated with the codes *unistructural*, *multistructural*, and *relational* of the SOLO Taxonomy. Of the 12 students taking part in the study, the covariational reasoning of six students was rated *unistructural*, the reasoning of three students was rated *multistructural*, and of three students it was rated *relational*. Fitzallen concluded that "the introduction of the concept of covariation can be adopted for upper primary students" (p. 240), but there were still shortcomings, since half of the students showed *unistructural* covariational reasoning.

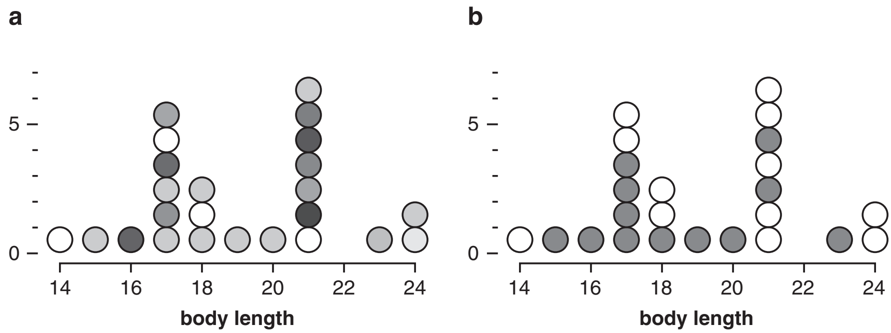


Fig. 5.12 Examples of using the color gradient in TinkerPlots for the variables body length and age (**a**) and for the variables body length and gender (**b**) similar to Konold (2002, pp. 4–5)

5.6.5 Reflections and Future Developments

Having a look back on Sect. 5.6, many different aspects of association and covariation have arisen. Research on learners' reasoning about covariation and association in different settings (e.g., contingency tables, scatterplots, scatterplots with superimposed lines or curves) and examples of software tools that might enhance learners' reasoning about association and covariation have been considered in this section.

One thing that consistently comes out in the research on covariation is the challenge of getting students to reason from the data itself and to consider setting aside previous beliefs. The whole paradigm of making data-based decisions and data-based inferences is quite foreign to many students. This is one finding that challenges both researchers and teachers to get students into situations and tasks, where they can confront their own biases and beliefs, and begin to put on a "data detective" hat.

5.7 Conclusion

This chapter has primarily concentrated on research on student and teacher reasoning about *variation*, *distribution*, *comparing groups*, and *associations and covariation between variables*. One main idea may be identified across this chapter: frameworks to rate students' and teachers' statistical reasoning. Many qualitative studies applied the SOLO Taxonomy to rate learners' statistical reasoning in responses to open-ended complex tasks, and the SOLO taxonomy has been used to categorize such reasoning for nearly all statistical concepts. While the SOLO Taxonomy proves to be a powerful tool for rating learner reasoning in different fields, future empirical studies might tend to consider more open analysis methods like Grounded Theory (Corbin & Strauss, 1994) or an inductive-enriched Qualitative Content Analysis (Mayring, 2015), ensuring that frameworks for student reasoning arise from the data (e.g., interviews, videos) itself.

With regard to technology, this chapter showed the potential of using appropriate software to enhance student and teacher reasoning about data. Educational software like Fathom and TinkerPlots can support student reasoning about distribution in varying displays of data and moving them from a local to a global perspective on data. When comparing groups, software may help students see differences between groups in even large data sets and enable learners to use individual comparison approaches (like p-based comparisons). When dealing with associations of variables and covariation, tools like Fathom and TinkerPlots can provide deeper insights when switching views between cell, column, and row percentages and help to produce sliced scatterplots so that covariational reasoning can be related back to students' understanding of univariate distributions and comparisons of groups. It is still an important issue in future research to better understand the purpose, the constraints, and the possibilities of digital tools and to research learning environments that make use of technologies.

Several recurring issues for the teaching and learning of data analysis have emerged throughout this chapter. The research has consistently found that students tend to provide surface level descriptions when asked to analyze data, interpret graphs, or compare groups. To better develop learners' reasoning about data, the research suggests that it may be important to provide learning activities that systematically guide them from a local to a global view on distributions, to sensitize learners to variability in the data, and to motivate them to go beyond surface level descriptions of what can be seen in data to a more substantial interpretation. For example, it is important to encourage learners to move beyond just reading the data to reading beyond the data (prediction, inference, and interpretation) and to reading behind the data (contextual and data production issues) and eventually to looking for relationships with other variables and to potential causes (for the three levels *reading the data*, *reading between the data*, and *reading behind the data*, see, e.g., Friel et al., 2001). This might be done by providing students with norms for adequate and non-adequate interpretations of data and graphs. Similarly, for developing learners, providing them with a data analysis scheme when reasoning about comparing groups may help them to better structure and to document their findings in their exploration process. Here it would be also important to provide learners with a process which can help them move from a mere description of findings to a deeper interpretation of findings when comparing groups. The development and implementation of structured norms for the analysis and interpretation of data may prove to be a fruitful area for future research on the teaching and learning of data analysis.

Reasoning about data was important in the past, is important in the present, and will be—especially in the upcoming big data era—even more critical for the future. Nowadays there is a huge quantity of data available via the Internet and other media, and many decisions in politics, economics, and society are based on statistics. In this respect, data science which combines disciplines like mathematics, statistics, computer science, and information science has the aim to cope with the huge amount of data and to extract important information of the data. Data science will be a rising and important field (of teaching and research) in the future.

So, to achieve informed participation in public decision processes, it is inevitably vital for concerned citizens to be statistically literate. To educate statistically literate citizens, teachers have to be prepared well for teaching statistics. Collaboration and intellectual conversations have been found to be a good way for teachers to improve their own reasoning about data. Teacher educators should include thoughtful attention to data analysis in the training and professional development of preservice and in-service teachers, especially since many teachers have little experience themselves in data analysis and statistical thinking.

For researchers, the future challenge is to deepen the research on learners' conceptions and misconceptions when reasoning about data. For teachers, the future challenge is to design and enhance activities and learning environments to develop sustainable reasoning about data across all ages of students.

References

Arnold, P. (2013). *Statistical investigative questions: An enquiry into posing and answering investigative questions from existing data* (Doctoral thesis). Retrieved from https://researchspace.auckland.ac.nz/handle/2292/21305.

Australian Curriculum Assessment and Reporting Authority. (n.d.). *Australian curriculum*. ACARA. Retrieved from http://www.australiancurriculum.edu.au

Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about box plots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education (IASE) roundtable* (pp. 163–173). Voorburg, The Netherlands: International Statistical Institute.

Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer.

Batanero, C., Estepa, A., & Godino, J. (1997). Evolution of students' understanding of statistical association in a computer-based teaching environment. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 198–212). Voorburg: International Statistical Institute.

Batanero, C., Estepa, A., Godino, J., & Green, D. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*(2), 151–169.

Batanero, C., Godino, J., & Estepa, A. (1998). *Building the meaning of statistical association through data analysis activities*, *Proceedings of PME-22, Stellenbosch, South Africa*. Alwyn: Olivier.

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal, 3*(2), 42–63.

Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics, 88*, 291–303.

Biehler, R. (1997). Students' difficulties in practicing computer supported data analysis—Some hypothetical generalizations from results of two exploratory studies. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 169–190). Voorburg: ISI.

Biehler, R. (2001). Statistische Kompetenz von Schülerinnen und Schülern—Konzepte und Ergebnisse empirischer Studien am Beispiel des Vergleichens empirischer Verteilungen [Statistical competence of students—Concepts and results of empirical studies on the example of comparing groups]. In M. Borovcnik, J. Engel, & D. Wickmann (Eds.), *Anregungen zum Stochastikunterricht* (pp. 97–114). Franzbecker: Hildesheim.

Biehler, R. (2004, July). *Variation, co-variation, and statistical group comparison: Some results from epistemological and empirical research on technology supported statistics education*. Paper presented at the Tenth International Congress on Mathematics Education, Copenhagen.

Biehler, R. (2005, February). Strength and weaknesses in students' project work in exploratory data analysis. In M. Bosch (Eds.), *Proceedings of the Fourth Congress of the European Society for Research in Mathematics Education, Sant Feliu de Guíxols, Spain* (pp. 580–590). Retrieved from http://ermeweb.free.fr/CERME4/CERME4_WG5.pdf.

Biehler, R. (2007a). Denken in Verteilungen—Vergleichen von Verteilungen [Thinking in distributions—Comparing distributions]. *Der Mathematikunterricht, 53*(3), 3–11.

Biehler, R. (2007b). *Students' strategies of comparing distributions in an exploratory data analysis context*. CD-ROM Proceedings of 56th Session of the International Statistical Institute. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/isi56/IPM37_Biehler.pdf

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (Vol. 27, pp. 643–689). New York: Springer.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Biggs, J. B., & Collis, K. (1991). Multimodal learning and the quality of intelligent behavior. In H. Rowe (Ed.), *Intelligence, reconceptualization and measurement* (pp. 57–76). NJ: Laurence Erlbaum Associates.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education—A joint ICMI/IASE study: The 18th ICMI study* (pp. 57–69). Dordrecht: Springer.

Canada, D. (2006). Elementary pre-service teachers' conceptions of variation in a probability context. *Statistics Education Research Journal, 5*(1), 36–64.

Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: A framework and a study. *Journal for Research in Mathematics Education, 33*(5), 352–378.

Casey, S. A., & Nagle, C. (2016). Students' use of slope conceptualizations when reasoning about the line of best fit. *Educational Studies in Mathematics, 92*(2), 163–177.

Ciancetta, M. (2007). *Statistics students reasoning when comparing distributions of data* (Unpublished doctoral dissertation). Portland State University, Portland, OR.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*(1), 1–78.

Common Core State Standards Initiative. (2015). Retrieved from http://www.corestandards.org/

Contreras, J. M., Batanero, C., Diaz, C., & Fernandes, J. A. (2011). Prospective teachers' common and specialized knowledge in a probability task. In M. Pytlak, T. Rowland, & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp. 766–775). Rzeszów: University of Rzeszów, Poland.

Corbin, J., & Strauss, A. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273–285). Thousand Oaks, CA: Sage.

Engel, J., & Sedlmeier, P. (2011). Correlation and regression in the training of teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education—A joint ICMI/IASE study: The 18th ICMI study* (pp. 247–258). Dordrecht, The Netherlands: Springer.

Estrada Roca, A., & Batanero, C. D. (2006). Computing probabilities from two way tables: An exploratory study with future teachers. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics: Working cooperatively in statistics education, Salvador, Brazil [CD-ROM]*. Voorburg, The Netherlands: International Association for Statistical Education and the International Statistical Institute.

Finzer, W. (2007). *Fathom dynamic data software* [computer software]. Emeryville, CA: Key Curriculum Press. [Software is available from www.concord.org]

Fitzallen, N. (2012). *Reasoning about covariation with TinkerPlots* (Unpublished Ph.D. Thesis). University of Tasmania, Australia.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K–12 curriculum framework*. Alexandria, VA: American Statistical Association.

Friel, S., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124–158.

Friel, S. N., O'Connor, W., & Mamer, J. D. (2006). More than "meanmedianmode" and a bar graph: What's needed to have a statistical conversation. In G. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance: Sixty-eighth NCTM yearbook* (pp. 117–137). Reston, VA: National Council of Teachers of Mathematics.

Frischemeier, D. (2014, July). Comparing groups by using TinkerPlots as part of a data analysis task—Tertiary students' strategies and difficulties. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, AZ.* Voorburg, The Netherlands: International Statistical Institute.

Frischemeier, D. (2017). *Statistisch denken und forschen lernen mit der Software TinkerPlots*. Wiesbaden: Springer Spektrum.

Frischemeier, D., & Biehler, R. (2016). Preservice teachers' statistical reasoning when comparing groups facilitated by software. In K. Krainer & N. Vondrova (Eds.), *Proceedings of the 9th Congress of the European Society for Research in Mathematics Education* (pp. 643–650). Faculty of Education and ERME: Charles University in Prague.

Gal, I., Rothschild, K., & Wagner, D. A. (1989). *Which group is better? The development of statistical reasoning in elementary school children*. Paper presented at the meeting of the Society for Research in Child Development, Kansas City.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal, 4*(1), 92–99.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and practice*. The Netherlands: Springer.

Gigerenzer, G. (1996). The psychology of good judgment frequency formats and simple algorithms. *Medical Decision Making, 16*(3), 273–280.

Gigerenzer, G. (2002). *Reckoning with risk: Learning to live with uncertainty*. London, UK: Penguin.

Graham, A. (1987). *Statistical investigations in the secondary school*. Cambridge, UK: Cambridge University Press.

Hasemann, K., & Mirwald, E. (2012). Daten, Häufigkeit und Wahrscheinlichkeit. In G. Walther, M. van den Heuvel-Panhuizen, D. Granzer, & O. Köller (Eds.), *Bildungsstandards für die Grundschule: Mathematik konkret* (pp. 141–161). Berlin: Cornelsen Scriptor.

Hogan, T. P., Zaboski, B. A., & Perry, T. R. (2015). College students' interpretation of research reports on group differences: the tall-tale effect. *Statistics Education Research Journal, 14*(1), 90–111.

Inhelder, B., & Piaget, J. (1955). *De la logique de l'enfant á la logique de l'adolescent [The growth of logical thinking from childhood to adolescence]*. Paris: Presses Universitaires de France.

Jones, G. A., Langrall, C. W., Mooney, E. S., & Thornton, C. A. (2004). Models of development in statistical reasoning. In J. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201–226). Dordrecht, The Netherlands: Kluwer.

Kader, G. D., & Perry, M. (1994). Learning statistics. *Mathematics Teaching in the Middle School, 1*(2), 130–136.

KMK. (2004). *Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss*. München: Wolters Kluwer.

KMK. (2012). *Bildungsstandards im Fach Mathematik für die allgemeine Hochschulreife*. Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf.

Konold, C. (2002). Alternatives to scatterplots. *Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. *A research companion to principles and standards for school mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

Konold, C., & Miller, C. (2011). *TinkerPlots TM Version 2 [computer software]*. Emeryville, CA: Key Curriculum Press. [Software available from www.tinkerplots.com]

Konold, C., & Pollatsek, A. (2002). Data analysis as a search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: 1996 Proceedings of the 1996 IASE Round Table Conference* (pp. 169–190). Voorburg, The Netherlands: International Statistical Institute.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., et al. (2002). *Students' use of modal clumps to summarize data. Paper presented at the Sixth International Conference on Teaching Statistics*. South Africa: Cape Town.

Lane, A. (2015). Simulations of the distribution of the mean do not necessarily mislead and can facilitate learning. *Journal of Statistics Education, 23*(2).

Langrall, C., Nisbet, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing groups. *Mathematical Thinking and Learning, 13*(1–2), 47–67.

Lehrer, R., Kim, M., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modeling variability. *International Journal of Computers for Mathematical Learning, 12*(3), 195–216.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 635–679.

Lem, S., Kempen, G., Ceulemans, E., Onghena, P., Verschaffel, L., & Van Dooren, W. (2014, July). Teaching box plots: An intervention using refutational text and multiple external representations. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, AZ*. Voorburg, The Netherlands: International Statistical Institute.

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2013). External representations for data distributions: In search of cognitive fit. *Statistics Education Research Journal, 12*(1), 4–19.

Madden, S. R. (2008). *High school mathematics teachers' evolving understanding of comparing distributions* (Unpublished Dissertation). Western Michigan University.

Makar, K. (2004). *Developing statistical inquiry: Prospective secondary math and science teachers' investigations of equity and fairness through analysis of accountability data* (Ph.D. Thesis). University of Texas at Austin.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal inferential inference. *Mathematical Thinking and Learning, 13*(1–2), 152–173.

Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 1–4.

Makar, K., & Confrey, J. (2002). *Comparing two distributions: Investigating secondary teachers' statistical thinking*. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.

Makar, K., & Confrey, J. (2004). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 353–373). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27–54.

Makar, K., & Confrey, J. (2014). Wondering, wandering or unwavering? Learners' statistical investigations with Fathom. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen—Using tools for learning mathematics and statistics*. Springer Spektrum: Wiesbaden.

Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. In A. Bikaner-Ahsbahs, C. Knipping, & N. C. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 365–380). Dordrecht: Springer.

McKenzie, C. R., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology, 54*(1), 33–61.

Meletiou-Mavrotheris, M., & Lee, C. (2002). Teaching students the stochastic nature of statistical concepts in an introductory statistics course. *Statistics Education Research Journal, 1*(2), 22–37.

Ministry of Education. (2007). *The New Zealand curriculum.* Ministry of Education, New Zealand Government. Retrieved from http://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20–39.

Mooney, E., Duni, D., VanMeenen, E., & Langrall, C. (2014). Preservice teachers' awareness of variability. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS-9), Flagstaff, AZ.* Voorburg, The Netherlands: International Statistical Institute.

Moritz, J. (2004). Reasoning about covariation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227–256). Dordrecht, The Netherlands: Kluwer Academic Publishers.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Noll, J. (2011). Graduate teaching assistants' statistical content knowledge of sampling. *Statistics Education Research Journal, 10*(2), 48–74.

Noll, J., & Shaughnessy, M. (2012). Aspects of students' reasoning about variation in empirical sampling distributions. *Journal for Research in Mathematics Education, 43*(5), 509–556.

Noss, R., Hoyles, C., & Pozzi, S. (2002). Abstraction in expertise: A study of nurses' conceptions of concentration. *Journal for Research in Mathematics Education, 33*(3), 204–229.

Noss, R., Pozzi, S., & Hoyles, C. (1999). Touching epistemologies: Meanings of average and variation in nursing practice. *Educational Studies in Mathematics, 40*(1), 25–51.

Obersteiner, A., Bernhard, M., & Reiss, K. (2015). Primary school children's strategies in solving contingency table problems: the role of intuition and inhibition. *ZDM, 47*(5), 825–836.

Pérez Echevarría, M. P. (1990). *Psicología del razonamiento probabilístico [The psychology of probabilistic reasoning]*. Madrid: Ediciones de la Universitad Autónoma Madrid.

Peters, S. (2011). Robust understanding of statistical variation. *Statistics Education Research Journal, 10*(1), 52–88.

Peters, S. (2014). Developing understanding of statistical variation: Secondary statistics teachers' perceptions and recollections of learning factors. *Journal of Mathematics Teacher Education, 17*(6), 539–582.

Pfannkuch, M. (2005). Thinking tools and variation. *Statistics Education Research Journal, 4*(1), 83–91.

Pfannkuch, M. (2007). Year 11 Students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education, 2*(3), 149–167.

Pfannkuch, M., & Budgett, S. (2017). Reasoning from an eikosogram: An exploratory study. *International Journal of Research in Undergraduate Mathematics Education, 3*(2), 283–310. https://doi.org/10.1007/s40753-016-0043-0

Pfannkuch, M., Budgett, S., & Parsonage, R. (2004). *Comparison of data plots: Building a pedagogical framework.* Paper presented at the Tenth Meeting of the International Congress on Mathematics Education, Copenhagen, Denmark.

Pfannkuch, M., & Reading, C. (2006). Reasoning about distribution: A complex process. *Statistics Education Research Journal, 5*(2), 4–9.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, thinking and reasoning* (pp. 17–46). Dordrecht, The Netherlands: Kluwer Academic Press.

Reaburn, R. (2012). Strategies used by students to compare two data sets. In J. Dindyal, L. P. Cheng, & S. F. Ng (Eds.), *Mathematics education: Expanding horizons (Proceedings of the 35th annual conference of the Mathematics Education Research Group of Australasia).* MERGA: Singapore.

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal, 3*(2), 84–105.

Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal, 5*(2), 46–68.

Reading, C., & Reid, J. (2007). Reasoning about variation: Student voice. *International Electronic Journal of Mathematics Education, 2*(3), 111–127.

Reading, C., & Reid, J. (2010). Reasoning about variation: Rethinking theoretical frameworks to inform practice. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistics Institute.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201–226). Dordrecht, The Netherlands: Kluwer.

Reid, J., & Reading, C. (2006). A hierarchy of tertiary students' consideration of variation. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS-7), Salvador, Brazil*. International Statistics Institute: Voorburg, The Netherlands.

Reid, J., & Reading, C. (2008). Measuring the development of students' consideration of variation. *Statistics Education Research Journal, 7*(1), 40–59.

Reid, J., & Reading, C. (2010). Developing a framework for reasoning about explained and unexplained variation. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistics Institute.

Rossman, A., & Chance, B. (2001). *Workshop statistics: Discovery with data* (2nd ed.). Emeryville, CA: Key College Publishing.

Rossman, A. J., & Chance, B. L. (2014) Using simulation-based inference for learning introductory statistics. Wiley Interdisciplinary Reviews: Computational Statistics 6 (4):211–221. Doi:10.1002/wics.1302

Rubin, A., Bruce, B., & Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. Vol. 1, pp. 314–319). Voorburg, The Netherlands: International Statistical Institute.

Saldanha, L., & Thompson, P. (2003). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257–270.

Sanchez, E., Borim da Silva, C., & Coutinho, C. (2011). Teachers' understanding of variation. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 211–221). Dordrecht, The Netherlands: Springer.

Schnell, S., & Büscher, C. (2015). Individual concepts of students comparing distributions. In K. Krainer & N. Vondrova (Eds.), *Proceedings of the 9th Congress of the European Society for Research in Mathematics Education* (pp. 754–760). Prague: Charles University in Prague and the European Society for Research in Mathematics Education (ERME).

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester & NCTM (Eds.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Information Age Publications: Charlotte, NC.

Shaughnessy, J. M., Chance, B., & Kranendonk, H. (2009). *Focus in high school mathematics. Reasoning and sense making: Statistics and probability*. National Council of Teachers of Mathematics. Reston: VA.

Shaughnessy, J. M., Ciancetta, M., & Canada, D. (2004). Types of student reasoning on sampling tasks. In M. Johnsen Høines & A. Berit Fuglestad (Eds.), *Proceedings of the 28th meeting of the International Group for Psychology and Mathematics Education* (Vol. 4, pp. 177–184). Bergen, Norway: Bergen University College Press.

Slauson, L. V. (2008). *Students' conceptual understanding of variability* (Unpublished PhD Dissertation). The Ohio State University.

Taylor, L., & Doehler, K. (2015). Reinforcing sampling distributions through a randomization-based activity for introduction ANOVA. *Journal of Statistics Education, 23*(3).

Vergnaud, G. (1996). The theory of conceptual fields. In L. P. Steffe, P. Nesher, P. Cobb, G. Goldin, & B. Greer (Eds.), *Theories of mathematical learning* (pp. 219–239). Hillsdale, NJ: Erlbaum.

Wassner, C. (2004). *Förderung Bayesianischen Denkens: kognitionspsychologische Grundlagen und didaktische Analysen; mit Arbeitsmaterialien und didaktischen Kommentaren zum Thema "Authentisches Bewerten und Urteilen unter Unsicherheit" für den Stochastikunterricht der Sekundarstufe I.* Hildesheim [u.a.]: Franzbecker.

Watkins, A., Bargagliotti, A., & Franklin, C. (2014). Simulation of the sampling distribution of the mean can mislead. *Journal of Statistics Education, 22*(3).

Watson, J. M. (2009). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal, 8*(1), 32–61.

Watson, J., & Callingham, R. (2014). Two-way tables: Issues at the heart of statistics and probability for students and teachers. *Mathematical Thinking and Learning, 16*(4), 254–284.

Watson, J., & Callingham, R. (2015). Lung disease, indigestion, and two-way tables. *Investigations in Mathematics Learning, 8*(2), 1–16.

Watson, J., Callingham, R., & Donne, J. (2008). Proportional reasoning: Student knowledge and teachers' pedagogical content knowledge. In M. Goos, R. Brown, & K. Makar (Eds.), *Navigating currents and charting directions. Proceedings of the 31st Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 563–571). Brisbane, Australia: Mathematics Education Research Group of Australasia.

Watson, J. M., Callingham, R. A., & Kelly, B. A. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical Thinking and Learning, 9*(3), 83–130.

Watson, J., Fitzallen, N., Wilson, K., & Creed, J. (2008). The representational value of HATS. *Mathematics Teaching in Middle School, 14*(1), 4–10.

Watson, J. M., & Kelly, B. A. (2004a). Expectation versus variation: Students' decision making in a chance environment. *Canadian Journal of Science, Mathematics and Technology Education, 4*(3), 371–396.

Watson, J. M., & Kelly, B. A. (2004b). Statistical variation in a chance setting: A two-year study. *Educational Studies in Mathematics, 57*(1), 121–144.

Watson, J. M., & Kelly, B. A. (2006). Expectation versus variation: Students' decision making in a sampling environment. *Canadian Journal of Science, Mathematics and Technology Education, 6*(2), 145–166.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*(1), 1–29.

Watson, J. M., & Moritz, J. B. (1999). The beginnings of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*, 145–168.

Watson, J. M., & Moritz, J. B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2*(1), 11–50.

Watson, J. M., & Shaughnessy, J. M. (2004). Proportional reasoning: Lessons from research in data and chance. *Mathematics Teaching in the Middle School, 10*(1), 104–109.

Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal, 5*(2), 10–26.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(1), 223–265.

Zieffler, A. S., & Garfield, J. B. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal, 8*(1), 7–31.

Zieffler, A., Harring, J., & Long, J. D. (2011). *Comparing groups: Randomization and bootstrap methods using R*. Hoboken, NJ: Wiley.

Chapter 6
Research on Uncertainty

Dave Pratt and Sibel Kazak

Abstract We discuss research on the teaching and learning of uncertainty, with a particular emphasis on quantifiable aspects as might be represented by probability. We acknowledge earlier reviews of the field by integrating research, especially from the last 10 years, with previous studies. In particular, we focus on three issues, which have become increasingly significant: (1) the realignment of previous work on heuristics and biases, (2) conceptual and experiential engagement with uncertainty and (3) adopting a modelling perspective on probability. The role of the teacher in shaping the learning environment in various critical ways emerges as a key finding. In the concluding section, we indicate promising directions for research, including the need for more exploratory research in new areas such as the role of modelling and carefully designed experiments to test hypotheses that are apparent from more established studies.

Keywords Probability • Heuristics • Biases • System 1 • System 2 • Modelling • Scaffolding • Dialogic thinking • Distribution • Sample space

6.1 Introduction

The notion of uncertainty is a broad concept that includes phenomena that lie outside the domain of statistics, which focusses on uncertainty due to random variation, when it is often possible to make inferences and predictions. Within this subset of uncertainty, it is sometimes possible to measure how uncertain a phenomenon is, and we refer to this term as 'probability'. Probability theory provides tools for expressing, quantifying and modelling uncertainty. This chapter focusses on research concerning those key ideas and issues in uncertainty and probability that

D. Pratt (✉)
Institute of Education, University College London, London, UK
e-mail: david.pratt@ucl.ac.uk

S. Kazak
Faculty of Education, Pamukkale University, Denizli, Turkey
e-mail: skazak@pau.edu.tr

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_6

are seen as conceptual links to statistics. We first give an overview of the previous reviews related to the topic and then introduce our approach to reviewing the research literature beyond those covered in the preceding documents.

There have been several edited books (e.g.Chernoff & Sriraman, 2014; Jones, 2005; Kapadia & Borovcnik, 1991) and a number of major review chapters and reports on research in probability since probability and statistics started to become part of the mainstream school mathematics curricula in many countries. In his review in the *Handbook of Research on Mathematics Teaching and Learning*, Shaughnessy (1992) set the stage by addressing the absence of probability and statistics in school mathematics, particularly in the USA prior to *The Curriculum and Evaluation Standards for School Mathematics* (NCTM, (1989). He then used philosophical and historical influences in the development of probability as the backdrop for research in probability and statistics. In Shaughnessy's (1992) review, studies in the research literature were clustered in three main areas: (1) different types of thinking used in making an inference or judgement under uncertainty (i.e. heuristics, biases and misconceptions) that are identified and documented primarily within the psychology research tradition, such as the influential work of Daniel Kahneman and Amos Tversky in the 1970s and 1980s, (2) development of concepts of probability in different age levels and (3) effects of interventions, such as types of tasks, instructional approaches and use of computer technology, on students' conceptions of probability.

Another important review of the existing research at that time by Borovcnik and Peard (1996) focussed on probabilistic thinking and teaching of probability in the school mathematics curriculum. Borovcnik and Peard shed light on what hindered learning of probability by making distinctions between probability and other mathematical concepts and similarly between probabilistic thinking and other types of thinking (logical and causal). They also described how the history of teaching probability evolved both in Europe and the USA as probability and statistics became part of the school mathematics curricula in different countries. Then various didactical approaches aiming to enhance teaching probability were discussed.

In the *Second Handbook of Research on Mathematics Teaching and Learning*, Jones, Langrall, and Mooney's (2007) chapter revealed how much progress had been made both in the treatment of probability in curriculum documents and the research tradition since Shaughnessy's (1992) review. One of the focusses of Jones et al. (2007) was the content and pedagogical insights of three curriculum documents from the USA, the UK and Australia at different grade levels (elementary, middle and high school), which were published around the same time. In these curriculum documents, introduction of probability to students started at early grades in the elementary school while they began to focus on more advanced ideas in probability at the high school level. In terms of pedagogy, the tasks at the elementary grades were more in line with students' experiences in the way that they allowed students to test their intuitions and overcome their misconceptions. As students reached advanced levels, investigations and applications were emphasised. Simulations and modelling chance situations, which had been suggested in Shaughnessy's (1992) review, were seen to become part of the middle grades and high school curricula (Jones et al., 2007).

The research literature covered in Jones et al. (2007) also reflected these changes in mathematics curricula by focussing on various conceptual topics relevant to probability, such as chance and randomness, sample space, probability measurement (including conditional, theoretical and empirical probabilities) and cognitive models of probabilistic reasoning. Related to the research on teaching of probability, Jones et al. highlighted the contributions made about teachers' content knowledge, pedagogical content knowledge and knowledge of student cognition in probability. Another distinct topic raised in this review was the idea of probability literacy based on Gal's (2005) work and its implications for content and pedagogical approaches in probability instruction.

Bryant and Nunes (2012) provided a detailed report on research documenting children's difficulties in learning and reasoning about probability and recommendations for future research, particularly on methodological aspects. They argued that four ideas in probability were key to successful learning in probability: (1) understanding randomness and its consequences, (2) analysing the sample space, (3) quantifying probability as a ratio and (4) developing correlational reasoning which involved the coordination of the previous three ideas. The omission of aggregate thinking as relating to distribution—rather than just sample space—is surprising in the light of research reported below.

More recently Watson, Jones, and Pratt (2013) took a critical approach when reviewing research studies into students' reasoning about uncertainty, many of which were mentioned in the previous reviews. Unlike others written mainly for researchers, the primary aim of this work was to elaborate the research-based findings to support pre-service and in-service teachers' understanding of the key issues about students' learning about probability. Given the technological tools that have become available in recent years, the use of simulations and modelling to help students develop reasoning about uncertainty was again emphasised in Watson et al.'s review, including for young students.

Our aim in this chapter is to focus on three issues that we see as key developments emerging out of the history of the topic as captured by previous reviews mentioned above. The first issue, the realignment of heuristics and biases, is chosen because the research on heuristics has been a major focus of research in the field for many decades and a recent publication makes it timely to reconsider that body of work. The second issue, conceptual and experiential engagement with uncertainty, gives an account of recent developments in the main effort of research in the field, some of which might in fact be influenced by the first issue. The third issue, adopting a modelling perspective on probability, emerges directly from considerable development in the use of technology for teaching and learning and also for researching students' ideas about uncertainty. As we introduce each of these three key issues in the following sections, we give an overview of the research in learning and teaching of probability over the last 10 years and look forward to future research in this domain.

6.2 The Realignment of Heuristics and Biases

6.2.1 Introduction

We begin our review with discussion of an issue that has informed—some would say beset—research on probabilistic thinking for several decades. The issue in question was a particular focus of the review of research in probability and statistics, which is now more than two decades old (Shaughnessy, 1992), and has led to an industry of research identifying misconceptions and correlating them in support of, or in contradiction to, the original work. We speak of course about the seminal work by Daniel Kahneman and Amos Tversky (e.g. Kahneman, Slovic, & Tversky, 1982), which claimed to catalogue the biases inherent in heuristics that we all use to make judgements of chance. This research has recently achieved new currency because of Kahneman's publication on *Thinking, Fast and Slow* (Kahneman, 2011a), which has reconceptualised the original research and, in doing so, has responded to some of the original criticisms. Kahneman's realignment of his own work on heuristics has implications for interpretation of the wealth of research on probabilistic thinking, especially as related to misconceptions, that has emanated from the original work in the 1970s and 1980s.

Our approach to discussing this key issue will be first to summarise the original work. This can be done briefly since there are many full accounts available elsewhere, not least in the review by Shaughnessy (1992). We will then discuss some of the criticisms that emerged over subsequent years. All of this will be preparation for a detailed account of Kahneman's fresh perspective on that work, followed by a discussion of whether the old criticisms still stand and implications for research in the field.

Kahneman and Tversky conducted a series of carefully designed psychological experiments where subjects were given tasks, either orally or in paper and pencil form. Kahneman and Tversky noted the errors that subjects made when their responses were compared with normative probabilistic or statistical solutions to the task. They identified a number of patterns in those errors and accounted for these patterns in terms of the subjects using rules of thumb, perhaps subconsciously, which they referred to as heuristics. Kahneman and Tversky explained how errors resulted from the bias, which was inherent in the heuristics being used.

As explained above, it would be inappropriate here to detail the huge array of heuristics identified, especially as each of the heuristics also has many variations and specific types. Nevertheless, some readers may wish to have a sense of the original work without needing to know all of that detail, so we will describe here two of the main heuristics identified by Kahneman and Tversky.

6.2.2 *Two Heuristics from the Original Research and Recent Developments*

When people use the representativeness heuristic, they judge the likelihood of an event according to how well the outcome experienced matches the system that generated the outcome or the population from which the outcome was drawn. The well-known gambler's fallacy might be accounted for by the use of the representativeness heuristic. Thus, after observing six successive red numbers appear on the roulette wheel, the gambler might place his bet on the appearance of a black number (an approach referred to as negative recency). Kahneman and Tversky argue that the gambler might believe that the outcomes should match the sample space, which consists of an equal number of red and black numbers, and so the judgement may have been made that a black number should appear in order to 'correct' the sequence of reds. The representativeness heuristic operates in the gambler's judgement by attempting to match the outcomes with the sample space.

Another situation in which the representativeness heuristic can lead to erroneous judgement is when a specific condition is regarded as more probable than a single general one, often referred to as the conjunction fallacy. For example, given a pen portrait description of Linda as a woman who is single, outspoken and very bright and deeply concerned with issues of discrimination and social justice, it is not unusual for subjects to respond that Linda is more probably a bank teller and active in the feminist movement than that Linda is a bank teller.

Kahneman and Tversky argue that the representativeness heuristic will often provide correct judgements, but since representativeness does not allow for the vagaries of random chance, nor the laws of probability, there will be situations in which representativeness generates the wrong judgement, a systematic error that the authors refer to as 'bias'.

A second major heuristic identified by Kahneman and Tversky is 'availability'. People sometimes make a judgement about the chance of an event on the basis of how easily they are able to evoke particular instances of the same or similar events. For example, the risk of a crash by the aeroplane in which you are travelling may seem disproportionately high (when compared to the frequency of recorded accidents) if there has been a recent widely reported tragic case of such an incident in which many people died. As with representativeness, the availability heuristic will often generate a correct judgement, but how easily instances of an event can be evoked is highly sensitive to the salience of the event. The salience of an event is not generally related to its likelihood, which results in a bias inherent in the availability heuristic.

In the last 10 years, there has been further research on the trajectory over time of heuristics and biases and errors. Bennett (2014) studied 163 first year college students (though this group was divided into several treatments, so the sample size for any one experiment was in the low 30s). The study found that students working on tasks inspired by the Monty Hall problem demonstrated a strong tendency for their decisions to be shaped by the 'endowment effect', an unwillingness to tempt fate by changing one's mind about a decision in the light of further information, even when a rational decision according to probability theory would be to do so.

Chiesi and Primi (2009) investigated how the errors due to negative (and positive) recency developed or receded with age. They tested 23 primary school third graders, 25 primary school fifth graders and 35 college students. They found that whereas positive recency (e.g. in which the gambler would bet on another red number at the roulette wheel after a sequence of red numbers) decreased with age, the negative recency effect was unaffected over time.

Kustos and Zelkowski (2013) examined misconceptions in probability tasks in the form of a survey consisting of open-ended structured questions for between 500 and 600 students across grades 7, 9, 11 and also those of 40 third year pre-service mathematics teachers. These misconceptions included inter alia recency effects and representativeness, in other words some of the errors that arise, according to Kahneman and Tversky, from biases in heuristic thinking. They found that the recency effect and representativeness dissipated with age.

There is an evident discrepancy between the above two studies of how misconceptions arising from heuristic thinking develop. The large-scale study of Kustos and Zelkowski suggests that factors in the development of students in middle- to high-achieving schools in Alabama have a positive impact on the students' probabilistic reasoning. Although the researchers offer implications for teaching, these must be regarded as speculative as pedagogy was not investigated in the study. The smaller study of Chiesi and Primi took place in Italian public schools, and it is entirely possible that factors impacting on the development of these students were very different. It is also possible that the sample size in this study was too small. Further research is needed before we can understand these conflicting results, and it may be that a better theoretical understanding of heuristic thinking is needed before we can really predict how errors might be affected by schooling or age. A new theoretical understanding is perhaps now beginning to emerge and is discussed later in this section.

6.2.3 *Criticisms of Kahneman and Tversky's Original Work on Heuristics*

The main critic of the heuristics and biases approach of Kahneman and Tversky has been Gerd Gigerenzer. In his own work, Gigerenzer has advocated the use of natural frequencies instead of probabilities or proportions to communicate risk (e.g. Meder & Gigerenzer, 2014). Bodemer, Meder, and Gigerenzer (2014) demonstrated that people with a wide range of numeracy levels were less likely to interpret relative risk reductions in heart disease as absolute reductions when the baseline risks were presented in frequency format than when they were presented as percentages. We note however that Diaz and Batanero (2009) conducted a comparison of performance amongst 206 students, who took a test after a teaching unit on conditional probability, with a comparable group of 177 students, who took the test before the course. They argue that detailed analysis of the types of errors apparent at different stages of a solution led them to teaching approaches that have demonstrated improvement in performance, even when probabilities rather than natural frequencies were used in conditional probability problems. (Specific cases where they did not find improvement are reported below.)

Gigerenzer (1991) has argued that some of the errors identified by Kahneman and Tversky disappear when the information is presented in a frequency format. Kahneman (1996) in turn has responded that their own studies supported the notion that presentation format impacted on the use of heuristics. However, he argued that this did not undermine the observation that subjects made systematic errors when presentations were not frequency based. Kahneman added that, though these errors might have been reduced, they did not disappear when the format was changed to one of natural frequencies, except perhaps in some very specific types of heuristic, such as the conjunction fallacy. Interestingly, in the Diaz and Batanero (2009) study, using probabilities rather than natural frequencies, the conjunction fallacy was one of the few errors that was resistant to improvement through their teaching methods.

Gigerenzer (1994) has also argued that there are difficulties with Kahneman and Tversky's focus on errors, which requires a normative position against which to judge the subjects' responses. They argue that there is fundamental disagreement amongst statisticians about the nature of probability, especially in relation to unique events, where a frequentist interpretation of probability does not apply. Of course, in many situations frequentist and subjective interpretations of probability converge and Kahneman (1996) pointed out that much of their historical work was not based around subjective probabilities. In fact, Gigerenzer's (1993) philosophical position regarded people's use of heuristics as rational acts, where decision-making apparatus has evolved so that decisions can be made when time and resources are limited. In his view, such apparatus rationally seeks out heuristic-based methods of decision-making at the expense of accuracy and that these rational methods can be more accurate than formal methods. Hence, whereas Kahneman and Tversky have presented a fallible human who makes errors due to the use of inherently biased heuristics, Gigerenzer has offered a rational human who uses heuristics that are often accurate to make quick decisions on complex judgements of chance.

Perhaps, this theoretical difference goes to the heart of the prolonged dispute that has stretched across many publications, through the 1980s and 1990s. Although in his review, Shaughnessy (1992, p. 470) referred to the Kahneman and Tversky work as providing a theoretical framework for mathematics educators, one criticism of the work has been that it is in fact atheoretical. In response, Kahneman (1991) argued:

> I take the distinctive feature of theory to be a commitment to completeness (within reason) and a consequent commitment to critical testing, in a specified domain of refutation, which is often quite narrow. (p. 143)

The difficulty for educationists lies in how to interpret Kahneman and Tversky's original work without a theoretical account of knowledge, thinking and learning. The catalogue of errors might be interpreted as suggesting that fallibility with respect to judgements of chance is integral to the human condition, which would be a bleak interpretation for those who hope to intervene in a student's understanding of probability. On the other hand, perhaps awareness of the bias in the use of heuristics, such as representativeness and availability, could be sensitised with the possible effect of improved judgements of chance. Clearly, as psychologists with an interest in decision-making, Kahneman and Tversky were not attempting to offer advice to educationists. Nevertheless, the recent publication, *Thinking, Fast and Slow*, by Kahneman (2011a), does situate the original research in a theoretical framework, which makes it possible to interpret the implications of the original work in new ways, and perhaps sheds new light on the theoretical difference between Kahneman and Gigerenzer.

6.2.4 Heuristics as Part of System 1 and System 2 Thinking

Kahneman (2011a) has recently adopted dual process theory, and in particular the terminology of Stanovich and West (2000), to refer to System 1 thinking as automatic, quick and requiring little or no effort with no sense of voluntary control. In contrast, he stated that System 2 thinking is effortful, often involving complex computations, associated with agency, choice and concentration. To take one of Kahneman's examples, look at the following problem: 17 × 24. System 1 tells you immediately that this is a multiplication problem (and may even allow you to estimate a rough answer). However, to compute the actual value requires the slow thinking of System 2. Loosely speaking, if System 1 were regarded as intuition, System 2 could be thought of as formal reasoning. Kahneman argued that much decision-making, and certainly that which involved the heuristics he had identified in his earlier work, operated at the automatic, largely subconscious level of System 1, whereas the careful application of scientific theory and procedures to reach a decision demanded the effort of System 2. While System 1 by default was triggered automatically to make quick decisions, often with limited evidence, occasionally System 2 was activated when System 1 ran into trouble, such as when System 1 did not generate an answer, but System 2 required more time and resources. We can offer another illustrative example taken from a probability study by Kazak (2015). Consider a game in which two bags contain counters. One bag has three blue counters and one red counter. The other has one blue and three red counters. A player chooses one counter from each bag and wins if the colours match. Typically, the symmetry of the bags leads students to a swift judgement (System 1) to think the game is fair insofar as there appears to be an equal chance of winning. However, a careful calculation of the sample space (System 2) shows that the chances of winning and losing are not equal.

It is worth noting that Fischbein (first in his seminal work of 1975 and then through many subsequent publications) developed a substantial account of the part played in probabilistic thinking by our primary (unschooled) and secondary (systematically trained) intuitions. The description of System 1 thinking seems to match rather well this account of primary and secondary intuitions.

It is interesting to note that Babai, Brecher, Stavy, and Tirosh (2006), studying probabilistic reasoning, reported results that could be interpreted as supporting the operation of System 1 and System 2 in that reasoning. They studied the responses and response times of 68 16- and 17-year-old Israeli students to 20 'congruent' test items where the solution was expected to be in line with an intuitive response and 20 'incongruent' items in which the solution was regarded as counter-intuitive. They found not only that accurate responses were more prevalent amongst congruent items but also that correct responses to congruent items were quicker than correct responses to incongruent items. This finding is consistent with System 1 finding immediate solutions to the congruent items but System 2 needing to find a more effortful solution to the incongruent items.

System 1 cannot be switched off (Stanovich & West, 2000), so training System 2 to be less accepting of System 1 when System 1 readily finds a solution may become the focus for educationists. In Fischbein's terms, this could be one focus for promoting secondary intuitions. This is especially important for teachers and researchers of probabilistic thinking, identified by Kahneman as a conceptual field where System 1 often uses non-stochastic intuitions to respond to uncertainty.

At one level, we might recognise here Gigerenzer's portrayal of an evolved system that allows the heuristics (of System 1) to operate much of the time, perhaps forgoing accuracy (of System 2), for the benefit of speed (of System 1). Kahneman identifies a rich host of mechanisms that System 1 uses in order to reach quick answers to questions. For example, he claims that one technique is to substitute an easier question than the one actually posed. According to Kahneman, substitution is in fact a particularly prevalent cause of heuristic errors in the field of probability and statistics. For example, System 1 cannot correlate information about baseline frequencies alongside intuitions about resemblance, and so the representativeness heuristic tends to determine the decision. According to Kahneman, faced with a question about likelihood, System 1 substitutes a simpler question about resemblance. Another example is apparent when System 1 substitutes a question about the frequency of an event with a question about how easily similar instances come to mind, with the consequence that the availability heuristic tends to determine the answer.

Chernoff (2012) demonstrated the use of attribution substitution in probabilistic reasoning amongst 59 pre-service elementary and middle school teachers. In an unusual variation on the tasks typically used to test for representativeness, the subjects were asked which of the two answer keys (A C C B D C A A D B or C C C B B B B B B B) was least likely to be the answer key for a ten-question multiple choice math quiz, each question having four possible responses. They were also asked for an explanation. (An answer key is the coded list of correct responses.) Chernoff concluded that certain individuals, when presented with one question, possibly unknowingly answered a different question, substituting a variety of heuristic attributes, such as 'most resembling' in place of 'most likely'.

In contrast to Gigerenzer's emphasis on the rationality of people's use of heuristics, Kahneman's focus remains on how people's reliance on System 1 leads to systematic errors.

6.2.5 Implications of System 1 and System 2 for Probabilistic Thinking

One of System 1's techniques for making quick decisions is to readily draw causal inferences from the evidence immediately available. When presented with data, System 1 will begin to observe patterns and form impressions as possible causal explanations. System 2 typically accepts these explanations. This accounts for how we mistakenly see patterns in random behaviour, design in arbitrary events and intention in the accidental. According to Kahneman, this technique of System 1 explains why people, when presented with randomly generated data, use heuristics to predict how the sequence will extend. This attribute of System 1 also provides an account of why people confuse association with causation, attributing causality to patterns in data that might have no causal connection in reality:

> People are prone to apply causal thinking inappropriately to situations that require statistical reasoning. Statistical thinking derives conclusions about individual cases from properties of categories and ensembles. Unfortunately, System 1 does not have the capability for this mode of reasoning; System 2 can learn to think statistically, but few people receive the necessary training. (Kahneman, 2011a, p. 77)

In the Diaz and Batanero (2009) study, participants often confused causality and conditionality, and they typically assumed that the likelihood of an event could not be affected by the likelihood of an event that has already happened. These errors were resistant to improvement through their teaching methods. Perhaps because System 1 searches for causations, there is a tendency to account for conditional relationships as if they were causal with time dependence.

The difficulty people have in recognising a situation as amenable to a statistical interpretation has been well documented. Konold (1989) referred to people's tendency to focus on what happened, rather than on strategic probabilistic approaches, as the 'outcome' approach. Thus, focussing on outcomes, System 1 might easily infer causations even when the patterns noticed are explained merely by the vagaries of chance.

Lecoutre, Rovira, Lecoutre, and Poitevineau (2006) investigated how 20 grade 3 pupils, 20 psychology researchers and 20 mathematics researchers, all based in Rouen in France, decided on whether given situations might involve randomness.

Sixteen items were presented on cards and varied according to whether: (1) the items were events from everyday life experiences or a repeatable process that might involve random variation; (2) the items addressed the subject as 'you' or not; (3) the possible outcomes were equally likely or asymmetric. First, the subjects were asked to categorise the 16 items for themselves, and then they were asked which items involved randomness. The researchers concluded that subject-decided randomness was involved when they could recognise probabilistic reasoning, for example, by being able to compute a probability, making probability rather than randomness the foundational idea. Subject-decided randomness was not involved when they thought determinism played the larger part or when causal factors could be identified. Since System 1, according to Kahneman, is constantly searching for causal patterns, it is perhaps not surprising that the possibilities for a stochastic approach tend to be ignored and people demonstrate the outcome approach.

Smith and Hjalmarson (2013) examined 32 pre-service mathematics teachers' conceptions of random processes with respect to the traditional game of 'rock, paper, scissors'. Teachers found it difficult to reconcile equality of winning outcomes for each player with the human interference apparent when choosing how to place their fingers in the game. System 1 all too easily recognises the human element as a causation, but this conflicts with notions of fairness, often associated with randomness (Paparistodemou, 2014; Paparistodemou, Noss, & Pratt, 2008; Pratt, 2000). The pre-service teachers did ultimately decide that the outcomes were not generated randomly. The researchers concluded that understanding of the nature of randomness developed during their instructional sequence as a result of making the generating process explicit and focussing on whether that constituted random generation or not.

In their early work, Kahneman and Tversky introduced the so-called law of small numbers to describe how people behave as if the law of large numbers applies to short sequences as well. There is a tendency to underestimate the need for samples with large numbers in order to draw reliable inferences. Kahneman now explains this in terms of System 1. When samples are small, apparent patterns can be identified simply because extreme results are more likely to happen than if the sample were large, and System 1 tends to attribute causal explanations to those patterns.

More broadly, Kahneman argues that we easily think associatively, metaphorically and causally and these styles of thinking are more suited to System 1 than is statistical thinking. It remains an open question as to whether educationists will be able to find ways of training or educating their students such that System 2 would be less accepting of System 1 answers in identifiable scenarios. There is some evidence, presented below, to suggest that this may be possible. At the point that System 2 is required to affirm System 1's answer, it might be possible to teach System 2 to be less easily convinced in certain scenarios that capture typical probabilistic and statistical situations.

6.2.6 Intervention Studies on Heuristics and Biases

Below, we focus on intervention studies, which may suggest pedagogic methods to address the difficulties that seem to be generated through System 1 thinking.

Fast (2007) conducted a study of 54 female Zimbabwean students. A test, consisting of questions of the sort used by previous researchers to identify misconceptions, was administered. The students were found to make errors in their responses that were consistent with representativeness, availability and other heuristics. Source analogues were constructed and offered to the students through interviews. These analogues were designed to be structurally similar to the initial test items but were intended to generate normative responses and so be the basis of knowledge reconstruction, which was evaluated as generally successful. For example, a source analogue might pose a similar problem to that posed by the original test item but with the situation amended so that the numbers were more extreme. Thus, in the original test, subjects were asked whether a sports team, thought to be better performer, would be more likely to win against a supposedly inferior team in a playoff based on five matches or nine matches. In contrast, the analogue question compared a single playoff match with a five-game playoff. The intention was that subjects would be able to use common sense to find the correct response to the analogue question and then recognise its structural similarity with the original test item. A delayed post-test suggested that the analogues continued to provide anchors for normative thinking 1 month later. The process of knowledge reconstruction was seen as critical. Even though this research was based on a fairly small and specific group, the above intervention raises the question whether, in Kahneman's terminology, the use of analogues might offer a bridge towards normative thinking by sensitising System 2 to a set of scenarios in which System 1's automatic and quick response might otherwise be problematic.

Another approach has been demonstrated over several years, in the work of Pratt (2000) and Pratt and Noss (2002, 2010), where the intervention was based around children mending computer-based 'gadgets', virtual simulations of everyday random generators, whose configuration could be edited to make them work properly. These 10- to 11-year-old children tended not to recognise that attributes of randomness in the short term (e.g. unpredictability, lack of control over the outcomes, irregularity in results) differed from attributes of randomness in the long term, at least from the aggregated perspective (where relative frequencies become predictable and aggregated results have a regularity to them). From the Kahneman perspective, these children's System 1 heuristic thinking appeared to suggest that, when chance was operating, it was just a matter of luck. By working with the gadgets, the children gradually became aware of patterns in the aggregated view over the long term. Pratt and Noss (2010) concluded that the key elements in the intervention design were (1) enabling the testing by children of their personal conjectures, (2) seeking to enhance the explanatory power of knowledge that might offer a route to normalised knowledge, (3) constructing a task design that would be seen by the children as purposeful and allow them to appreciate the power of the mathematical idea of distribution and (4) designing a representation of distribution that could be initially used as a control point by the children and subsequently become a representation with predictive power. These design constructs perhaps offer some further insight into what might be needed in order to sensitise System 2 to the need to distinguish between scenarios with small and those with large numbers.

Paparistodemou et al. (2008) also used a computer-based microworld to study twenty-three 5- to 8-year-old children's ideas about fairness. The children were challenged to build a lottery machine by arranging a spatial configuration of red and blue balls, of which a small white ball would bounce. When the white ball hit a red ball, a character called the 'space kid' moved in one direction, and when a blue ball was hit, the space kid moved in the opposite direction. The aim was to keep the space kid near to his starting position. Some of the children's configurations exploited symmetry so that in effect the white ball bounced in turns from red to blue and back to red. Others exploited random bouncing so that it was impossible to predict which colour would be hit next. These two approaches were associated with deterministic and stochastic strategies, respectively. By placing emphasis on fairness in an expressive environment, the children were able to imagine fairness not only in terms of turn taking but also in terms of the vagaries of chance. The design constructs listed above (Pratt & Noss, 2010) seem to apply to this study as well, especially with respect to 1, 2 and 3. Kahneman might argue that the approach used in the Paparistodemou et al. (2008) provides System 2 with new possibilities for how fairness, when detected by System 1, might be interpreted.

An intervention by Canada (2006) might be seen as analogous to that by Paparistodemou et al. but with respect to variation in probability situations. Canada's use of hands-on activities, supplemented by small-group and whole-class discussion of variation, with pre-service teachers may enhance their appreciation of how variation plays a role in statistical thinking.

Another approach that might enhance students' System 2 recognition of the possible weakness in System 1's proposed solution is to improve teachers' pedagogical knowledge of the types of reasoning students might use. Such a development might alert teachers to the need to artificially engage their students' System 2 thinking, with the aspiration that, after sufficient training, their students might begin to recognise such situations for themselves. There appears at least to be a deficit in teachers' knowledge about students' probabilistic reasoning. In an interesting study, Watson and Callingham (2013) examined the probabilistic reasoning of 247 students, mostly from years 7 to 11, and compared that to how their 26 teachers recognised their students' reasoning. Some of the students' reasoning was unfamiliar to the teachers suggesting that there might be value in finding ways of enhancing the teachers' pedagogical knowledge in this area.

6.2.7 Discussion

In this section, we have considered a key issue that has emerged in research on heuristics for making judgements of chance because of Kahneman's (2011a) recent publication on two reasoning systems. Our perspective is that this issue is very important for researchers in statistics education, who are interested in randomness and probabilistic thinking, because dual process theory allows us to interpret research in the field in new ways.

The debate between Kahneman and Gigerenzer continues. In *Thinking, Fast and Slow*, there are several references by Kahneman to Gigerenzer's criticisms. In fact, Kahneman takes the opportunity to criticise Gigerenzer's notion of fast and frugal heuristics on the basis that, in Kahneman's view, there is no imperative for the brain with its massive processing power to be frugal. Meanwhile, Gigerenzer (2012) has described how methods of making rational choices are inefficient when key factors influencing the decision are unknown. For more recent developments in this ongoing debate, see Kahneman (2011b) and Gigerenzer (2014), where there is a chapter on revolutionising schools through a risk-based curriculum. This emphasis on a risk-based curriculum is in line with Fischbein and other researchers who have argued for many years that the curriculum is predominately anchored in deterministic reasoning (deduction, proof, algorithms) and has historically ignored stochastic reasoning under uncertainty (statistical thinking).

Overall, we have summarised Kahneman's application of dual process theory to his research, and we have reinterpreted recent research in those terms as a means to offer insight into its implications. Nevertheless, we acknowledge that it is perhaps too early to offer a critical evaluation of the realignment of the heuristics research as proposed by Kahneman beyond the discussion above about implications. In subsequent sections, we address other issues which we see as recent key developments in research on probabilistic thinking, and although the emphasis will move away from Kahneman's *Thinking, Fast and Slow*, we invite the reader to attempt to interpret this research from that perspective, which might indeed yield further insights.

6.3 Conceptual and Experiential Engagement with Uncertainty

6.3.1 Introduction

Probability is a means to quantify uncertainty in random processes. Understanding how the concept of probability historically developed provides a perspective for interpreting current research results on students' conceptions of probability. One important aspect of probability that appeared in the mid-1600s is its duality (Hacking, 1975; Weisburg, 2014). The dual notion of probability implies that on the one hand probability is considered as degree of belief (subjective notion), and on the other hand it refers to stable frequencies in the long run (objective notion). Another approach to estimating probability, especially in games of chance, involves a priori method that requires an assumption of equiprobability.

Accordingly, there are three main schools of thought in probability theory that have different conceptions/interpretations of probability. From the classical view, the probability of an event is a ratio of the number of favoured outcomes to the total number of equally likely outcomes. In the frequentist view, the probability of an event is defined as the limit of the relative frequency of the observed outcomes as the number of trials increases indefinitely when a random experiment is repeated under identical conditions. The subjective interpretation of probability emphasises personal probability relative to our background knowledge and beliefs.

The ongoing historical debates about different interpretations of probability have been also reflected in school curricula and in teaching of probability, such as theoretical, empirical and subjective probabilities (see Jones et al., 2007). While existing research on heuristics revealed the inconsistencies between students' informal conceptions of probability and formal theory of probability (see earlier section on heuristic thinking), many recent research studies investigated how students' probabilistic conceptions developed and the ways to support them. In this section we focus on this body of research. The first part focusses on research that is primarily about students' understanding, though we suggest implications for teaching. Subsequent parts consider how such understandings might be influenced by teachers, through the tasks they choose, their pedagogic approaches and the tools they offer to their students.

6.3.2 Recent Research on Conceptual Development

Given the historical development of various meanings of probability, the concept of probability has a slippery aspect. Furthermore, the seminal works by Piaget and Inhelder (1951) and Fischbein (1975) offered a starting point for much research, reviewed in detail elsewhere (Borovcnik & Peard, 1996; Shaughnessy, 1992), that showed how the learning of probability is troublesome. More recently, several researchers have been particularly interested in the development of these conceptions from a variety of theoretical perspectives. Below we first summarise that work, and then, in the final subsection, we draw together the implications for teaching.

Kafoussi's (2004) study focussed on the early development of quantitative reasoning about the likelihood of chance events during a classroom teaching experiment in a kindergarten. Individual interviews with children were conducted before and after the teaching experiment. Responses of the 5-year-old children during the pre-interviews tended to rely on subjective beliefs when judging the likelihood of given events. While children were able to identify all possible outcomes of a single-stage chance experiment, they could not give a complete answer for a two-stage experiment. They also seemed to have difficulties in comparing the likelihood of events when the task involved comparing of numbers of objects in a box rather than sizes of sections on a spinner. The post-interview results suggested considerable progress in children's probabilistic thinking showing a shift from subjective conceptions to a 'naive quantitative reasoning' as in Jones, Langrall, Thornton and Mogill's framework (Jones, Langrall, Thornton, & Mogill, 1997, p. 121). Kafoussi argued that 5-year-olds' conceptual development was fostered during the teaching experiment as they began to (1) discuss what counted as 'different' outcomes in a two-stage experiment, (2) consider the empirical results from an experiment as a solution to a probability problem and (3) predict the results of a probability situation with equiprobable outcomes without conducting an actual experiment.

Prediger (2008) reported on a clinical interview study with ten pairs of 10–11-year-old children by focussing on their individual conceptions of chance situations in a game context before any probability instruction at school. Prediger found three categories of conceptions when children were explaining or justifying the outcomes or their predictions: everyday conceptions, empirical conceptions and theoretical conceptions. She was cautious about simply making a correspondence between these individual conceptions and three interpretations of probability (subjective, frequentist and classical). She suggested that some of these student conceptions could later be developed into a subjective conception of probability or a frequentist conception. However, one pair of students seemed to develop a notion of a classical interpretation of probability when talking about the number of different ways to find the sum of two die. Apart from this one example where the students had a learning trajectory progressing from everyday conceptions to the classical conception of probability, the other pairs seemed to move back and forth between different conceptions.

Prediger however did not treat the individual conceptions that were not theoretically sound as misconceptions in a traditional sense (i.e. (mis)conceptions to be substituted by the mathematically appropriate ones). Using the approach of horizontal development in the conceptual change research tradition, she considered students' everyday conceptions 'as concurrent conceptions which co-exist with newly developed mathematical conceptions even in the long run' (Prediger, 2008, p. 142). Similar to previous findings (Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993; Pratt & Noss, 2002), the students' fluctuations between different conceptions during the task suggested that an individual might hold a range of views (from informal to formal) at the same time and use different ones depending on how they perceived the stochastic situation or what they paid attention to (single outcome vs. long run or short-term vs. long-term contexts). The horizontal view suggested a complementary perspective to the vertical view of conceptual change focussing on transformation of misconceptions to mathematical conceptions. Adopting this approach to conceptual development in probability seemed to provide a valuable perspective on 'typical' persevering misconceptions and how to reconceptualise them to help learners.

Furthermore, Schnell and Prediger (2012) applied the vertical and horizontal conceptual change approach to the development of students' conceptions of the empirical law of large numbers. However, their main focus in this paper was on the theoretical contribution of their fine-grained method for analysing the microprocesses of constructing conceptions by using a notion of 'construct' as the unit of analysis and of building links amongst them as a webbing of constructs. By microprocesses, they referred to moving from an initial construct to an advanced one or changing the function of a construct as new relations between constructs were formed. Schnell and Prediger argued these microprocesses would contribute to the vertical and horizontal conceptual changes, suggesting the possibility of a successful trajectory from a 'haphazard' view of changes in the chance outcomes to a stabilised view of patterns in the long-term context.

As shown in previous research on heuristics, students often come to classrooms with alternative conceptions of probability. Teachers need to be aware of these different interpretations of probability for helping learners develop the formal ideas. From this perspective, the study of Liu and Thompson (2007), focussing on teachers' understandings of probability on various tasks, is of importance. Research was conducted with eight high school teachers participating in an 8-week seminar on teaching and learning of probability and statistics with deeper understanding from a constructivist perspective. Liu and Thompson focussed on teachers' 'stochastic conception of probability' which they aligned to the frequentist view; in contrast, they argued that a 'relative proportion conception of probability' can sometimes be drawn upon without consideration of a repeatable stochastic process. Some other non-stochastic interpretations of probability, observed in teachers' responses and discussions, seemed to resemble those that students often have, for example, (1) the outcome approach (Konold, 1989), (2) reduction of sample space for a probabilistic event (i.e. given that either an event will happen or it will not happen, the probability is either 1 or 0) and (3) the principle of indifference approach to probability (i.e. the probability is 50% because an event may happen or not). Liu and Thompson argued that these non-stochastic interpretations would actually depend on how people conceived the given situation.

6.3.3 The Impact of Task Design on Conceptual Understanding of Probability

Conceptual development of probabilistic ideas is, of course, shaped by experience. For example, according to Ainley, Pratt, and Hansen (2006), students' conceptual understanding of the utility of a probabilistic idea is connected with their sense of the purposefulness of the task in which they are engaged. In pedagogic situations, tasks set by the teacher can sometimes seem artificial, lacking purpose or relevance from the perspective of the student, perhaps because the teacher is very aware of their responsibility to teach the syllabus. The challenge, and it is recognised as non-trivial, is to create tasks that are seen as purposeful by the student but result in the student gaining appreciation of how the statistical idea is powerful in helping them to complete the task.

An example lies in Pratt's (2000) study of children configuring computer simulations of random generators such as coins, spinners and dice, referred to as gadgets. The children found the task of trying to make the gadgets work properly purposeful, and it led inexorably to them gaining a sense of how a probability distribution, contextualised in this study as the working box of the gadget, had the power to predict aggregated outcomes in the long term but not in the short term. More generally, Ainley et al. suggested a range of heuristics for designing tasks that are likely to connect purpose and utility; tasks might (1) have an explicit end product, (2) involve making something for another audience to use and (3) contain opportunities for pupils to make meaningful decisions.

6.3.4 Scaffolding and Dialogic Thinking

As seen in the previous sections, misconceptions or biases that hinder students' probabilistic thinking are well documented. There are a few research studies examining how the pedagogic approach of the teacher might facilitate learning of probability.

Corter and Zahner (2007) initially worked with 26 graduate students in an introductory statistics course to examine the use of external visual representations in probability problem-solving. Each participant was asked to solve eight probability problems using a structured interview protocol. This exploratory study indicated that students used a variety of visual representations and that the appropriate ones tended to facilitate students' problem-solving. Zahner and Corter (2010) further researched the role of the external visual representations on solving probability problems (such as what kinds of representations were used for different problems, how and when) with another 34 graduate students. The interview-based research suggested that certain representations used spontaneously by the students helped them perform better in solving particular problems compared to those not using any. Selecting and using appropriate external representations in presented problems seemed to be an important part of the problem-solving process in this study.

Ruthven and Hofmann (2013) described the development of a probability module for early secondary school using classroom-based design research. A distinctive feature of this module was its pedagogical approach that was based on prior research on effective ways of teaching mathematics and science, especially in the UK context. This pedagogical intervention involved a teaching approach where students were encouraged to express their ideas, give explicit reasons for their thinking and take different perspectives, an approach termed 'dialogic' (see Mercer & Sams, 2006). Dialogic talk used in small group work and whole class discussions during the activities became a tool that helped students move from their informal ideas about probability, including some of those heuristics and biases mentioned above (mainly used in System 1 thinking mode) to formal probabilistic reasoning (i.e. System 2). Further evidence from Kazak, Wegerif, and Fujita (2015a), working with groups of 10–12-year-old children, supported the idea that scaffolding for dialogue as well as for content, alongside the use of technological tools, helped to generate breakthroughs in probabilistic thinking.

Kazak, Wegerif, and Fujita (2015b) explored whether an analysis of two 12-year-old students' activity based on dialogic theory might offer new insights compared to a Piagetian or Vygotskian analysis. The students were exploring the fairness of a variety of chance games, which they played manually but also built in *TinkerPlots 2.0* software (Konold & Miller, 2011, http://www.tinkerplots.com/). The researchers found that the Piagetian and Vygotskian analyses ignored what for most viewers of the activity was a very obvious phenomenon. The recordings of the activity showed how the students engaged in laughter, sometimes quite raucous, a phenomenon ignored by Piaget and Vygotsky, but of great interest to Bakhtin, whose work inspired the dialogic approach (Bakhtin, 1986). According to the authors, laughter creates space and openness for participants to switch perspective and so to take the point of view of the other. More generally, they argued that switching perspective was facilitated by the good relationship between the participants, including the teacher, good humour being one indicator of such a relationship.

6.3.5 The Role of Technology

In considering how teachers might influence students' understanding, we have so far considered recent research on task design, scaffolding through external visualisations and dialogic approaches. We now consider the tools, in particular technological tools, that they might offer their students. Research continues to suggest that certain types of technology, used within carefully designed situations, can offer opportunities for probabilistic learning that stretch beyond those available in everyday experience. Biehler, Ben-Zvi, Bakker, and Makar (2013) provided a recent review on such possibilities at school level. That review emphasised in conclusion some recurrent important points in the design of the learning environment that incorporates the use of technology:

1. Skill was needed, by the user or the teacher, to know when it was appropriate to adopt a hands-on approach and when software might help.
2. One key feature of modern pedagogic statistical software lay in its dynamic, visual and personal nature.
3. One key focus needed to be on reasoning with aggregates.
4. The tension between adopting the power that technology offered and the time it took to learn and adapt to that technology needed to be addressed.

With our specific focus on probability, we elaborate below a few research-based studies which we believe add to the above list of specific proposals for the design of a probabilistic learning environment but which were not detailed in that broader review.

Earlier, we mentioned Pratt's (2000) study in which 10- and 11-year-old children began to acknowledge that there were regularities in the aggregated results of random processes even though the same could not be said in the short term. In the previous section on heuristics, we set out the design constructs that, according to Pratt and Noss (2010), supported the development of those insights. Apart from those aspects of the design, it is clear that the technological environment provided the opportunity to gather artificial experience of the long term because the technology offered systematic feedback, quickly and repeatedly, which would not usually have been the case in everyday experience.

Similar results have been reported by Lee and Lee (2009), when children cheered for a chosen colour to be the most frequent in repetitions of computer-simulated draws of marbles from a bag, only to find that the result was rather predictable, except in a short run. They concluded that, in similar conditions to those reported by Pratt and Noss, students began to notice variability in small samples and regularity in large samples. When it came to interpreting the impact of adding some new data in small samples (more change/instability) vs. in large samples (less change, more stability) in the computer simulation results, other semiotic tools, such as the use of metaphors in combination with technology (Abrahamson, Gutiérrez, & Baddorf, 2012), helped students make sense of the visual phenomenon.

Ben-Zvi, Aridor, Makar, and Bakker (2012) studied how children aged 10–11 years expressed uncertainty while they conducted informal investigations of data. The students used *TinkerPlots 2.0* to make informal inferences on samples of data where the sample size was gradually increased. The students initially oscillated between deterministic and relativistic statements. Eventually, a basic probabilistic language began to emerge. The authors concluded that more sophisticated inference-making was encouraged by attending to students' expressions of uncertainty when making judgements about trends in data.

Abrahamson, Berland, Shapiro, Unterman, and Wilensky (2006) proposed an additional role for the computer. The authors of the paper discovered conflicts in their interpretations of a computer simulation in which three boxes were randomly coloured green or blue. A single run resulted in any one of eight possible configurations, called keys (e.g. green, green, blue is one key). The authors happily ran the simulation without disagreement. When the authors began to create probabilistic models of the situation, they discovered their apparent agreement was not founded on the same epistemological assumptions. It was possible to model either the length of a run of repeated guesses until a specific key appeared or the frequency of a particular key in various size samples of guesses. The authors found it difficult to agree on how the first model failed to generate the expected bell-shaped curve, a disagreement that was only resolved when the authors had had the opportunity to programme the situations, were confident that the programme was bug-free and had corrected any errors in thinking through discussion. Programming on the computer was for them a necessary step to expose and critique underlying assumptions and models, differences, which had not been apparent from simply running a prepared simulation. Chaput, Girard, and Henry (2008) made a similar point about modelling, which has some commonalities with programming insofar as both require the learner to express their ideas about what is being programmed or modelled. They argued that the use of modelling in statistics education is a delicate process because of the problematic epistemological basis of probability. They contended that the advantage of using computers resides not so much in their power and efficiency as in the analysis of random situations that needs to be done in order to design the model and translate that design into computer instructions.

In a sense, programming and discussion in Abrahamson's reflective article above acted to bridge across the differing probabilistic assumptions that the authors had held. Abrahamson and Wilensky (2007) reported how the design of pedagogical situations, including the use of technology, supported students to bridge intuitively, cognitively or historically conflicting ideas in probability. They referred to these conflicting ideas as being at opposite poles of a learning axis. They set out to design bridging tools that were intentionally ambiguous with respect to these extremes. These tools were presented as part of a broader learning environment, designed to stimulate engagement with and argumentation about the epistemological ambiguity. There is a connection here in how Abrahamson and Wilensky exploited ambiguity to set up cognitive conflict, subsequently resolved through discussion, and how Pratt and Noss (2010) referred to blurring control and representation in the way that the computer-based simulations were configured and used.

In summary, we might ask what have we learned about the role of technology in the teaching and learning of probability to add to the findings in the Biehler et al. review (2013). Certainly there is support (Lee & Lee, 2009; Pratt, 2000) for the idea that extended experience with the virtual, repeatable and artificial experience offered by some technological environments can contribute to a focus on aggregate thinking called for in that review, with the result that students can begin to distinguish between variability in the short term and regularity in the long term. In addition, there is growing evidence (Abrahamson et al., 2006) that programming models might for some clarify epistemological distinctions in probability. Biehler et al. highlighted the concern that in some situations teachers might judge that adopting technological approaches is more time-consuming than is warranted by the benefits that accrue and this could be a view taken by some teachers with respect to programming. The development of bridging tools (Abrahamson & Wilensky, 2007) that have a degree of ambiguity with respect to contrasting epistemologies might offer a similar role to programming and be less time-consuming for the student.

6.3.6 Discussion

In the first section of this chapter, we summarised the research on heuristics and biases and reviewed recent developments in theory that linked that earlier work to System 1 and System 2 thinking. According to Kahneman's account, System 1 thinking is relatively automatic and is best controlled by careful training of System 2. In the current section, we set out to review recent research to build on earlier reviews about how that might best be done.

What is clear from this review is the critical role played by teachers. Examples of this, cross referenced to the literature drawn on in this section, are:

1. Offering more empirical hands-on experience of random variation (Biehler et al., 2013)
2. The artful selection of digital tools and other types of external representations (Pratt, 2000; Zahner and Corter, 2010; Lee & Lee, 2009, Biehler et al., 2013)
3. Focussing such experience on prediction to tease out what counts as different outcomes (Kafoussi, 2004)

4. Recognising the complexity of different epistemologies of probability and helping students to bridge the apparent discrepancies through programming or specially designed tools (Abrahamson et al., 2006; Abrahamson & Wilensky, 2007; Liu & Thompson, 2007; Prediger, 2008)
5. Acknowledging the importance of task design, since the situation in which random variation is met influences how people think about probability and because purposeful tasks can, if carefully designed, lead to a sense of the power of the probabilistic concepts (Ainley et al., 2006)
6. Offering opportunities for students to express their ideas with their peers and through technology so that ideas can be negotiated and perhaps converge (Ruthven and Hofmann, 2013; Ben-Zvi, Aridor, Makar, and Bakker, 2012, Kazak et al., 2015a, 2015b)

Some of the above ways in which teachers might support learning of probability are especially suited to an approach in which probability is seen as a key part of creating or exploring models of situations that are amenable to a statistical interpretation. Modelling is therefore the focus of the next section in this chapter.

6.4 Adopting a Modelling Perspective on Probability

6.4.1 Introduction

One of the striking developments in recent research on probability (and its connections to statistics more generally) is the increased emphasis on modelling. Models have always been a key element of statistics as a discipline in the way that they describe data probabilistically (e.g. in the form of probability distributions or analytical methods such as analysis of variance). According to Wild and Pfannkuch (1999), modelling is also an important component of statistical reasoning. The emergence of modelling in teaching and learning has no doubt been driven by the increasing access to technology and improved software, especially that aimed at learners. Modelling appears to have the potential to facilitate the methods by which teachers can support learners, as listed in the previous subsection. Indeed, modelling promises to offer a connection between data and probability (Konold & Kazak, 2008) that is meaningful to learners and may provide an approach that enables learners to appreciate the power of probability, at a time when dice and card games have become less of a focus of play for the younger generation than in the past.

Modelling approaches tend to place emphasis simultaneously on data and uncertainty. Models can be developed to fit real data, but the fit will not be exact, requiring a probabilistic element to the model in order to account for the variation in the data. Computational models can be executed to generate virtual data, which may approximately reflect the real data if the model was a good one.

Theoretical distributions and sample spaces can be thought of as models, and so we begin this section by considering research in these areas. Subsequently, we consider research that addresses explicitly how a modelling perspective on probability might influence understanding (see Chap. 7).

6.4.2 Understanding Empirical and Theoretical Distributions

In their earlier review of student learning of probability, Jones et al. (2007) commented that, in view of its importance in curricula, it was surprising that at that time there was little research on student conceptions of experimental probability. They did quote limited evidence about the difficulty students experience in making links between the sample space of a random generator and outcomes actually generated. They also noted the proclivity for students not to realise the connection with the use of large samples until they were able to spend extended periods working with simulations that allowed the use of samples of any size.

There has since then been further research on students' understanding of theoretical and empirical distributions.

Ireland and Watson (2009), researching 10–12-year-old students, concluded that it was insufficient for educators to focus on the calculation of theoretical probabilities and the observation of experimental outcomes. According to their study, the connection between experimental and theoretical probability needed to be taught and experienced explicitly, by encouraging the creation of new correct probabilistic intuitions, the prediction of outcomes, the performance of experiments and the evaluation of outcomes as advocated by Fischbein (1975).

More recently, English and Watson (2016) conducted such a teaching experiment on 91 9- and 10-year-olds, who tossed one and two coins, and explored relative frequencies through graphing in *TinkerPlots 2.0*, which they also used to simulate large-scale tossing of coins. They concluded that working with the sampler in *TinkerPlots 2.0* seemed to help students to recognise that the frequency of two heads and two tails approached 25% while the frequency of one head and one tail approached 50%. However, this experiment took place in only one school and on one school day.

It is commonly thought that students observe how data from an experiment converges on the theoretical distribution. In fact, Lee, Angotti, and Tarr (2010), reporting on how 11–12-year-olds used a computer simulation to decide which of six companies were producing fair dice, concluded that it was not the cycling between model and data that was critical but developing well-connected conceptual links between model and data. Konold et al. (2011) suggested that constructing such a link was non-trivial for some students who appeared to lack a notion of a 'true' probability. Their subject appeared to distrust the idea that the theoretical probability was in fact the true probability exactly because the theoretical probability almost always failed to predict exactly what happened when the experiment was repeated. Indeed, to them, it was the experimental probability that reported what really happened.

A teaching episode reported by Noll and Shaughnessy (2012) focussed on samples and sampling distributions in probability tasks. In this episode students were engaged in making inferences about both known and unknown mixtures of coloured objects (i.e. estimating population proportions) based on empirical data obtained from repeated sampling. Researchers studied the impact of team teaching between the regular teachers and the investigators across six middle and high school classrooms. They concluded that teaching which focussed explicitly on distributions, especially sample-to-sample variability, enhanced students' reasoning about empirical sampling distributions.

6.4.3 Understanding Sample Space

Bryant and Nunes (2012) conducted a literature review for the Nuffield Foundation on children's understanding of probability. They regarded working out the sample space as one of four key demands in learning about probability. Moreover, generating representations, such as tree diagrams, organised lists and dot plots, based on sample space outcomes can support drawing conclusions and provide evidence for predictions (Fielding-Wells, 2015; Kazak & Pratt, 2015). In their earlier review of student learning of probability, Jones et al. (2007) also noted the importance of sample space, but they reported a range of difficulties in a concept that was not as straightforward as might be thought. They quoted research that identified difficulties: (1) in identifying possible outcomes even in simple random experiments, (2) in systematically generating all outcomes and (3) through failing to consider the sample space when determining probabilities.

Nunes, Bryant, Evans, Gottardis, and Terlektsi (2014) reported on how to support generating and using the sample space in quantifying the probability of an event in primary grades. They claimed that the conceptual schemas, such as classification, logical multiplication and ratio, which children begin to develop earlier in other domains (i.e. subtraction), can be used in understanding sample space. Nunes et al. designed an intervention study to test their conjecture that sample space could be taught in primary school by building on children's prior knowledge of these three concepts. In their study, one group of 10–11-year-olds participated in a teaching programme focussing on classification, logical multiplication and the use of ratios to quantify the probability of an event. Another group of participants (a comparison group) received instruction promoting mathematical problem-solving that was not related to sample space and probability. The third group (a waiting list control group) was taught by the class teacher and did not participate in a particular teaching programme until after the study. The study showed that the children in the intervention programme performed significantly better than their counterparts in both comparison groups. However, there was no significant difference between the problem-solving group and the unseen control group on any of the post-tests. According to Nunes et al., an instructional programme promoting the use of tree diagrams supported students' development of combinatorial understanding. This in turn was needed to understand how to generate a sample space by building on the concepts of classification and logical multiplication. After systematically identifying all possible outcomes and classifying those into favourable and unfavourable cases in the sample space, students used the ratios to quantify the probability of an event. We note however that the suggested approach in this intervention study is only applicable to limited situations where the classical probability definition is used, where the sample space is discrete rather than continuous and where each possible outcome is equally likely.

The aggregation of cases (as favourable and unfavourable), mentioned by Nunes et al. (2014), is a crucial step in determining the probability of an event by using ratios. However, studies by Francisco and Maher (2005) and Nilsson (2007) indicate that this idea was challenging to students in complex probability situations. For example, Francisco and Maher's (2005) study showed that while students were able to list all possible outcomes in combinatorics problems, they had difficulties in identifying the sample space in a probability problem and, particularly, in determining the denominators of the probability ratio.

Nilsson (2007) focussed on the notion of sample space as a model for probability predictions in chance games. This study explored the strategies used by students (ages 12–13) when pairs were asked to distribute a set of markers on a game board numbered from 1 to 12 and to play the game against the other group by looking at the sum of two unusual dice. Students used the following pairs of designed dice in the game: (111222) and (111222), (222444) and (333555), (111122) and (111122) and (222244) and (333355), where, for example, (111222) represents a six-sided dice with three 1 s and three 2 s on its faces. In each of these four different game settings, an analysis of sample space for totals of two dice was required for making a decision about the distribution of markers on the game board. The study showed that students intuitively began to use what they considered as the sample space to decide the most/least likely totals in a given dice set-up. However, their focus was on the resulting sums by looking at only the proportions of numbers available on the individual dice rather than examining the number of different combinations to get each sum. Hence, their incomplete sample space provided a limited model for their decisions in different dice set-ups.

Abrahamson (2009a, 2009b) reported on the single case of Li, an 11-year-old student, using a specially designed scoop, which collected four marbles from a large pot, containing green and blue marbles in equal numbers. Any 1 scoop therefore contained 1 of 16 equally likely outcomes. First, Li was asked what would happen if the researcher were to scoop the marbles. Second, he was given card and crayons and asked to colour in all the different scoops. Third, Li was asked to create a combination tower, in effect a histogram of the number of (say) green marbles in a scoop. These tasks lay a foundation for the binomial probability distribution, which is typically one of the first formal models used by statisticians and taught in an advanced statistics course at high school level and in an introductory statistics course at university level. For example, they are relevant to modelling one-dimensional random walk problems, especially for young students (e.g. Kazak, 2010), and the distribution of gender in 12-children families (e.g. Biehler, Frischemeier, & Podworny, 2015).

In a detailed analysis of the clinical interview that took place around these three tasks, Abrahamson reported that Li's initial perception of the likelihood of events such as two green and two blue marbles was undermined by the need to construct the various permutations in the second and third tasks. Li saw no reason not to consider some of those permutations as redundant repeats. When the repeats were ignored, it seemed that there were five events (0, 1, 2, 3 and 4 blue marbles in a scoop) and there was no apparent reason for not thinking of these five as equally likely. According to Abrahamson, it was only when Li was able to make a 'semiotic leap' that he was able to use the tools to warrant his initial correct intuitive perceptions.

The use of such bridging tools might initially have been meaningless, but as the tools were well designed from a pedagogical and epistemological point of view, they led to semiotic leaps such as recognising why the events in the five-point sample space were not in fact equally likely. In the study reported by Pratt (2000), the students needed to realign fairness away from the totals of two dice to the individual combinations in what Abrahamson would have termed a semiotic leap.

Given the difficulties students often encounter in generating and using sample space in probability contexts, Chernoff and Zazkis (2011) suggested a new term, 'sample set', as a bridging tool between student-generated lists of outcomes and the conventional sample space consisting of equiprobable outcomes. A sample set was used to refer to any set of all possible outcomes of an event. For example, in Abrahamson's (2009a) four-marble task, {4 green and 0 blue, 3 green and 1 blue, 2 green and 2 blue, 1 green and 3 blue, 0 green and 4 blue} would be a sample set listing all possible outcomes of the scoop experiment. Unlike some students' thinking, this is not the sample space used in computing probabilities as ratios because the listed outcomes are not equiprobable. Consequently it leads to an incorrect answer as seen in Li's case (Abrahamson, 2009a). Chernoff and Zazkis argued for a pedagogical approach that 'without compromising mathematical rigour, acknowledges the learner and serves as a bridge between personal, sometimes naive, and conventional knowledge' (p. 19).

6.4.4 The Role of Modelling

For a typical statistician, a model can be imagined as a generator of data comprised of a main effect (signal) that explains much of the variation together with residual or unexplained variation, sometimes referred to as random error (Wild, 2006). With modern software, computational models can actually generate data, akin to the statistician's way of thinking about the model. In Sect. 6.3.1, we discussed the differing epistemologies of probability. Depending on the given situation, probability can be interpreted as a theoretical solution based on an equiprobable sample space, a relative frequency in the long run or a subjective degree of belief. Shaughnessy (1992) advocated a modelling perspective. As seen in several research studies in the Jones et al. (2007) review chapter, probability can be viewed as a tool for modelling uncertain situations and making simulation-based inferences (Watson, Jones, and Pratt, 2013).

Although several studies below have demonstrated some promise as to how a modelling approach might support aggregate thinking, in relating to distribution and sample space, learning to model is non-trivial. Indeed, speaking about science, Lehrer and Schauble (2010) emphasised the difficulties faced by novices. In fact, Pfannkuch and Ziedins (2014) proposed that more emphasis be placed on helping students to appreciate the purpose of modelling. More specifically they suggested that models be categorised as 'good' or 'bad' or otherwise that no model currently exists. In that way, they suggested that students could engage in modelling activity either to use a good model, improve a bad model or create a model where one does not exist.

6.4.4.1 The Role of Modelling in Understanding Distribution

Modelling promises to offer some leverage in dealing with the issues raised above about the challenge of connecting sample space, theoretical and empirical distribution. Konold, Harradine, and Kazak (2007) used a data modelling approach in exploring middle school students' understanding of distributions. The modelling activities in a series of tasks that focussed on a 'data factory' metaphor involved using *TinkerPlots 2.0* modelling capabilities to create a distribution that would match the expected data in the real world, such as hair length of females and males. Using a similar approach, Lehrer, Kim, and Schauble (2007) examined fifth–sixth grade students' use of *TinkerPlots 2.0* tools to model a distribution of repeated measurements of their teacher's head. Student-generated models included an estimate value of the true length of the circumference using the median of the real measurements and the combination of some random errors, such as reading error and ruler error. Comparing simulation results in *TinkerPlots 2.0* with the actual data helped students revise their model. Both studies suggested these types of data modelling tasks with young students as a foundation for important ideas in statistical inference.

Prodromou and Pratt (2006, 2013) studied pairs of students aged between 14 and 16 years as they worked with a specially designed microworld where the students controlled the throw of a basketball. Control was exerted through sliders, which controlled variables such as the angle of release. These variables worked either deterministically or stochastically by changing the parameter value and varying the spread around that value, thus introducing variation into the basketball throw. Within this setting, Prodromou and Pratt (2006) focussed on students' development of two perspectives on data generated by the computer simulations, which were called modelling and data-centric perspectives. They distinguished the two perspectives on distribution as they suggested different ways of perceiving variation. The researchers proposed that (1) the modelling perspective emerged when students manipulated the tools controlling the position and spread of the distribution and (2) the data-centric perspective was revealed when students focussed their attention on variation and the shape of the emerging data. They also argued that being able to coordinate these two perspectives was essential in viewing data as a combination of signal and noise, which is a fundamental idea in statistical thinking (Konold & Pollatsek, 2002).

Drawing upon the coordination of two perspectives on distribution, Prodromou's (2012) work with pre-service primary school teachers focussed on making connection between the empirical probability and the theoretical probability of the sum of two dice. The findings showed that pre-service teachers paid attention to the variation in the empirical data distribution (data-centric perspective on distribution) and the stability of the relative frequencies in the long run with a resemblance to the theoretical distribution (modelling perspective). A few of them also were able to make the connection from theoretical probabilities (modelling) to empirical probabilities (data-centric) as a way to make predictions.

6.4.4.2 The Role of Modelling in Understanding Sample Spaces

Konold and Kazak (2008) highlighted the model fit idea to connect the empirical distribution and the expected (theoretical) distribution. Within this approach students tried to make sense of observed data with regard to a model when making a prediction; they sometimes revised their model on the basis of data. Students tended to make their initial predictions based on their experiences or beliefs about the likelihood of random events, which were often in conflict with the accepted theory. Konold and Kazak argued that engaging students in developing the sample space in which the compound event occurred provided a theoretical model and facilitated their explanations for the distribution of actual and/or simulation data generated in *TinkerPlots 2.0*. They also suggested that by evaluating differences, or the fit, between the expected distribution based on the sample space and the distributions obtained from the simulations, they began to perceive observed data as a noisy version of the theoretical expectation (the signal) in relation to the size of data collected. Hence, this model fit approach provided a context to focus students' attention on sample space, which was often a challenging concept especially when students encountered compound events, as suggested by the studies mentioned in Sect. 6.4.3.

Most recently, the importance of the sample space analysis is also shown by the studies presented at the SRTL9, which investigated the role of building models in developing students' informal inference skills in games of chance (Fielding-Wells, 2015; Kazak & Pratt, 2015). In the context of a chance game seen on a popular television game show, Fielding-Wells (2015) discussed that structuring the sample space using a tree diagram provided a theoretical model and helped children (aged 10–11) make informal inferences based on the fit between the model and the data from experiments with the game device. In the context of another chance game involving the sum of two dice, Kazak and Pratt (2015) working with pre-service middle school mathematics teachers also reported on a case in which the probability model based on sample space emerged from engaging in both the combinatorial analysis of possible outcomes and empirical data both from playing the game physically and from simulations in *TinkerPlots 2.0*.

6.4.5 Discussion

Our review of research in this section suggests modelling as an emerging perspective for engaging students in probability contexts. This area of research is relatively new and still exploratory in the sense that conjectures are still being formed about how to support students' understanding of probability and ideas using a modelling approach.

As seen in the studies above, one advantage of the modelling perspective is that it brings statistical and probabilistic ideas together. These examples generally involve focussing on the match between the data generated empirically and the expected distribution based on sample space. In several of these studies, the role of technology is also worth noting in facilitating even very young students' understanding of probability. In addition, the modelling perspective appears to be relevant to promoting informal and formal statistical inference, which is addressed in Chap. 8 of this handbook, while students are expected to draw data-based conclusions. Research specifically on modelling is reported in Chap. 7.

6.5 Conclusion

In this final section, we summarise in broad terms each of the three central themes. For more detailed findings of our analysis, please refer to the discussions in each of the three main sections. In addition to this broad summary, we consider gaps in the research and future directions.

This chapter has focussed on research into how students learn to address uncertainty and how teachers support them in that process. The focus has been on that type of uncertainty that is more or less quantifiable. That is to say, we have not discussed research on somewhat less tangible aspects of uncertainty, such as the 'black swans' (Taleb, 2010), totally unpredictable events that can have dire consequences. While these other types of uncertainty are socially very important and interesting, the statistics educator is particularly concerned about situations that might incorporate randomness, quantified through probability. To this end, we have focussed here primarily on recent research, which we have contextualised within previous reviews of related research.

In the first section of this chapter, we discussed how the research on heuristics and biases has been represented as underpinned by dual process theory, potentially offering new insights into the many difficulties teachers and researchers have unearthed over the years regarding understanding probability. In particular, the new theoretical basis for the research on heuristics may point to innovative pedagogies to support the triggering of System 2 thinking when making judgements under uncertainty. We discussed some of the more promising research in this area. There needs to be further research to identify how, in Kahneman's terminology, System 2 might be better trained to recognise scenarios in which System 1's solution is likely to be biased. In Gigerenzer's terminology, research is needed to identify pedagogic approaches that lead to more accurate fast and frugal heuristics. The important theoretical distinction here is that Kahneman's ideas hold out little hope for improvement in System 1 but rather in identifying how better to use System 2, whereas Gigerenzer would focus on researching better heuristics.

In the second section, we elaborated further by considering the impact of how tasks are designed, how technology is adopted and more generally how students are taught on the development of probability as a concept. This research presents the clear conclusion that teachers are central if students are to develop the slow thinking of System 2 to manage in a more sophisticated way the quick intuitions of System 1.

The second section summarised how, post Jones et al. (2007), there has been an increasing number of research studies focussing on students' understanding of the relationship between experimental and theoretical probabilities with the availability of new technology tools. However, there is still a scarcity of research when the sample space is continuous and also in the area of subjective probability at the school level. We found no research on Bayesian methods at this level (see Chap. 13 for more on Bayesian methods). Pedagogical approaches, including task design, to bridging the three dominant interpretations of probability need to be developed and tested in classroom settings. The second section ends by summarising what appear to be key aspects of how teachers might have a positive effect. Further research on task, tool and activity setting design is needed to identify how best to offer hands-on purposeful experience that promotes discussion and prediction and bridges different epistemological perspectives.

The third section points out that, perhaps driven by advances in the use of technology and in software development for educational purposes, probability can be presented as a mathematical model of (quantifiable) uncertainty. Indeed, such software allows the student to express their understanding of chance in the form of computational probabilistic models that can be executed. A modelling perspective on probability seems to offer a bridge that might help learners to coordinate the potentially confusing classicist, frequentist and subjectivist epistemologies of probability.

At the very least, when students create such models, they engage in activity that crosses any artificial boundaries that may otherwise have been set up between probability and statistics. Curricula have for many years tended to separate probability from statistics. Such a separation might render probability somewhat meaningless as students struggle to recognise any utility for the topic. Modelling approaches can counter that danger. As well as the examples described in the third section above, there are many others scattered in the book as a whole (e.g. see Chap. 8 on informal statistical inference). Nevertheless, a modelling approach brings with it some new difficulties, touched on in the third section.

Most educational research on modelling in this field is recent because modern technological tools have opened up new possibilities; perhaps as a result, the promise that modelling offers to help learners link probability and statistics remains open to further exploration. There needs to be more exploratory research that clarifies how pedagogic approaches might exploit the potential of modelling for probabilistic learning while providing pathways through the obstacles for learners that no doubt will become more evident. One challenge is how to design tasks that make modelling seem purposeful to learners so that they can begin to engage with its utility or power. Another challenge is how to provide guidance on what makes a model effective. At the same time, there is still need for investigating the role of other visualisation tools (physical materials, diagrams and so on) and teacher scaffolding in promoting the modelling approach especially during off-computer tasks.

Although such research would be exploratory, there may be other research opportunities, which can test verifiable conjectures. Bryant and Nunes (2012) argue that much of the research on children's understanding of probability is based on good ideas but that its design is limited. They call for many more cross-sectional and longitudinal studies as well as intervention projects that test causal hypotheses about the factors involved in children's learning of probability. Testing causal hypotheses is difficult in educational research because there is an ethical dimension that resists the construction of randomised controlled trials. Nevertheless, there are now some examples of where this has been possible, and Bryant and Nunes call for more. The field is now relatively mature, and this review alongside earlier ones may help to identify opportunities for this type of systematic research that tests well-formulated hypotheses. Of course, there continues to be a need for exploratory studies in less well-developed topics, such as in the area of modelling, where clear and testable hypotheses are not yet available.

References

Abrahamson, D. (2009a). Orchestrating semiotic leaps from tacit to cultural quantitative reasoning—The case of anticipating experimental outcomes of a quasi-binomial random generator. *Cognition and Instruction, 27*(3), 175–224.

Abrahamson, D. (2009b). A student's synthesis of tacit and mathematical knowledge as a researcher's lens on bridging learning theory. In M. Borovcnik and R. Kapadia (Eds.), Research and developments in probability education [Special Issue]. *International Electronic Journal of Mathematics Education*, *4*(3), 195–226.

Abrahamson, D., Berland, M., Shapiro, B., Unterman, J., & Wilensky, U. (2006). Leveraging epistemological diversity through computer-based argumentation in the domain of probability. *Learning of Mathematics, 26*(3), 39–45.

Abrahamson, D., Gutiérrez, J. F., & Baddorf, A. K. (2012). Try to see it my way: The discursive function of idiosyncratic mathematical metaphor. *Mathematical Thinking and Learning, 14*(1), 55–80.

Abrahamson, D., & Wilensky, U. (2007). Learning axes and bridging tools in a technology-based design for statistics. *International Journal of Computers and Mathematical Learning, 12*(1), 23–55.

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal, 32*(1), 23–38.

Babai, R., Brecher, T., Stavy, S., & Tirosh, D. (2006). Intuitive interference in probabilistic reasoning. *International Journal of Science and Mathematics Education, 4*(4), 627–639.

Bakhtin, M. M. (1986). *Speech genres and other late essays (trans: McGee, V. W.)*. Austin, TX, University of Texas Press.

Bennett, D. (2014). Sticking to your guns: A flawed heuristic for probabilistic decision-making. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 261–281). Dordrecht: Springer Science+BusinessMedia.

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM: The International Journal on Mathematics Education, 44*(7), 913–925.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technologies for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel-Kreidt, & J. Kilpatrick (Eds.), *Third international handbook of mathematics education* (pp. 643–688). New York: Springer Science+Business Media.

Biehler, R., Frischemeier, D., & Podworny, S. (2015). Elementary preservice teachers' reasoning about modelling a "family factory" with TinkerPlots—A pilot study. *Proceedings of the International Collaboration for Research on Statistical Reasoning, Thinking and Learning, SRTL-9* (pp. 146–175). Paderborn, Germany.

Bodemer, N., Meder, B., & Gigerenzer, G. (2014). Communicating relative risk changes with baseline risk: Presentation format and numeracy matter. *Medical Decision Making, 34*(5), 615–627.

Borovcnik, M., & Peard, R. (1996). Probability. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 239–288). Dordrecht, The Netherlands: Kluwer.

Bryant, P., & Nunes, T. (2012). *Children's understanding of probability: A literature review*. London: Nuffield Foundation.

Canada, D. (2006). Elementary pre-service teachers' conceptions of variation in a probability context. *Statistics Education Research Journal, 5*(1), 36–64.

Chaput, B., Girard, J. C., & Henry, M. (2008). Modeling and simulations in statistics education. In C. Batanero, G. Burrill, C. Reading and A. Rossman (Eds.). *Joint ICMI/IASE Study: Teaching statistics in school mathematics. Challenges for teaching and teacher education.* Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.

Chernoff, E. J. (2012). Recognizing revisitation of the representativeness heuristic: An analysis of answer key attributes. *ZDM Mathematics Education, 44*(7), 941–952.

Chernoff, E. J., & Sriraman, B. (2014). *Probabilistic thinking: Presenting plural perspectives*. New York: Springer.

Chernoff, E. J., & Zazkis, R. (2011). From personal to conventional probabilities: From sample set to sample space. *Educational Studies in Mathematics, 77*(1), 15–33.

Chiesi, F., & Primi, C. (2009). Recency effects in primary-age children and college students. *International Electronic Journal of Mathematics Education, 4*(3), 259–274.

Corter, J. E., & Zahner, D. (2007). Use of external visual representations in probability problem solving. *Statistics Education Research Journal, 6*(1), 22–50.

Diaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education, 4*(3), 131–162.

English, L. D., & Watson, J. M. (2016). Development of probabilistic understanding in fourth grade. *Journal for Research in Mathematics Education, 47*(1), 28–62.

Fast, G. R. (2007). Analogical reconstruction of the probability knowledge of Zimbabwean female secondary school students. *Canadian Journal of Science, Mathematics and Technology Education, 7*(2–3), 149–181.

Fielding-Wells, J. (2015). Young children's development and use of models as evidence in support of informal statistical inference. *Proceedings of the International Collaboration for Research on Statistical Reasoning, Thinking and Learning, SRTL-9* (pp. 98–106). Paderborn, Germany.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht, Holland: Reidel.

Francisco, J. M., & Maher, C. A. (2005). Conditions for promoting reasoning in problem solving: Insights from a longitudinal study. *Journal of Mathematical Behavior, 24*(2–3), 361–372.

Gal, I. (2005). Towards 'probability literacy' for all citizens. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 39–63). New York: Springer.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond heuristics and biases. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. Volume 2, pp. 83–115). Chichester, UK: Wiley.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 284–313). London: Routledge.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129–161). Chichester, UK: Wiley.

Gigerenzer, G. (2012). Contribution to panel 'What Can Economists Know: Rethinking the Foundations of Economic Understanding'. *The Institute for New Economic Thinking's (INET) Paradigm Lost Conference*, Berlin. Retrieved June 5, 2017, from https://www.youtube.com/watch?v=DdEEwoKkfMA.

Gigerenzer, G. (2014). *Risk savvy: How to make good decisions*. London, UK: Allen Lane.

Hacking, I. (1975). The Emergence of Probability: A philosophical Study of Early Ideas about Probability, Induction and Statistical Inference. Cambridge: Cambridge University Press.

Ireland, S., & Watson, J. (2009). Building a connection between experimental and theoretical aspects of probability. *International Electronic Journal of Mathematics Education, 4*(3), 339–370.

Jones, G. (2005). *Exploring probability in school: Challenges for teaching and learning*. New York: Springer.

Jones, G. A., Langrall, C., & Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). Greenwich, CT: Information Age Publishing, Inc. and NCTM.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32*(2), 101–125.

Kafoussi, S. (2004). Can kindergarten children be successfully involved in probabilistic tasks? *Statistics Education Research Journal, 3*(1), 29–39.

Kahneman, D. (1996). On the reality of cognitive illusions. *Psychological Review, 103*(3), 582–591.

Kahneman, D. (2011a). *Thinking fast and slow*. Farrar, NY: Straus and Giroux.

Kahneman, D. (2011b). *Thinking fast and slow, Google Talks*. Retrieved June 5, 2017, from https://www.youtube.com/watch?v=CjVQJdIrDJ0.

Kahneman, D. (1991). Judgment and decision making: A personal view, Psychological Science, 2(3), 12–145.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Kapadia, R., & Borovcnik, M. (1991). *Chance encounters: Probability in education*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Kazak, S. (2010). Modeling random binomial rabbit hops. In R. Lesh, P. L. Galbraith, C. R. Haines, & A. Hurford (Eds.), *Modeling students' mathematical modeling competencies: ICTMA 13* (pp. 561–570). New York: Springer.

Kazak, S. (2015). 'How confident are you?': Supporting young students' reasoning about uncertainty in chance games through students' talk and computer simulations. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 29–55). Minneapolis, MN: Catalyst Press.

Kazak, S., & Pratt, D. (2015). Pre-service mathematics teachers' informal statistical inference when building probability models for a chance game. *Proceedings of the International*

Collaboration for Research on Statistical Reasoning, Thinking and Learning, SRTL-9 (pp. 138–145). Paderborn, Germany.

Kazak, S., Wegerif, R., & Fujita, T. (2015a). Combining scaffolding for content and scaffolding for dialogue to support conceptual breakthroughs in understanding probability. *ZDM Mathematics Education, 47*(7), 1269–1283.

Kazak, S., Wegerif, R., & Fujita, T. (2015b). The importance of dialogic processes to conceptual development in mathematics. *Educational Studies in Mathematics, 90*(2), 105–120.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*(1), 59–98.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning, 12*(3), 217–230.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1), Article 1. Retrieved from http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art1

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., et al. (2011). Conceptual challenges in coordinating theoretical and data-centered estimates of probability. *Mathematical Thinking and Learning, 13*(1–2), 68–86.

Konold, C., & Miller, C. D. (2011). *TinkerPlots 2.0: Dynamic data exploration*. Emeryville, CA: Key Curriculum.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*(5), 393–414.

Kustos, P., & Zelkowski, J. (2013). Grade-continuum trajectories of four known probabilistic misconceptions: What are students' perceptions of self-efficacy in completing probability tasks? *Journal of Mathematical Behaviour, 32*(3), 508–526.

Lecoutre, M. P., Rovira, K., Lecoutre, B., & Poitevineau, J. (2006). People's intuitions about randomness and probability: An empirical approach. *Statistics Education Research Journal, 5*(1), 20–35.

Lee, H. S., Angotti, R. L., & Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal, 9*(1), 68–96.

Lee, H. S., & Lee, J. T. (2009). Reasoning about probabilistic phenomena: Lessons learned and applied in software design. Technology Innovations in Statistics Education, 3(2). Retrieved from http://escholarship.org/uc/item/1b54h9s9.

Lehrer, R., Kim, M. J., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modeling variability. *International Journal of Computers for Mathematical Learning, 12*(3), 195–216.

Lehrer, R., & Schauble, L. (2010). What kind of explanation is a model? In M. K. Stein & L. Kucan (Eds.), *Instructional explanation in the disciplines* (pp. 9–22). New York: Springer.

Liu, Y., & Thompson, P. W. (2007). Teachers' understandings of probability. *Cognition and Instruction, 25*(2), 113–160.

Meder, B., & Gigerenzer, G. (2014). Statistical thinking: No one left behind. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 127–148). Dordrecht: Springer Science+Business Media.

Mercer, N., & Sams, C. (2006). Teaching children how to use language to solve maths problems. *Language and Education, 20*(6), 507–528.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston: National Council of Teachers of Mathematics.

Nilsson, P. (2007). Different ways in which students handle chance encounters in the explorative setting of a dice game. *Educational Studies in Mathematics, 66*(3), 293–315.

Noll, J., & Shaughnessy, M. (2012). Aspects of students' reasoning about variation in empirical sampling distributions. *Journal for Research in Mathematics Education, 43*(5), 509–556.

Nunes, T., Bryant, P., Evans, D., Gottardis, L., & Terlektsi, M.-E. (2014). The cognitive demands of understanding the sample space. *ZDM Mathematics Education, 46*(3), 437–448.

Paparistodemou, E. (2014). Children's constructions of a sample space with respect to the Law of Large Numbers. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic Thinking: Presenting plural perspectives* (pp. 600–611). Dordrecht: Springer Science+Business Media.

Paparistodemou, E., Noss, R., & Pratt, D. (2008). The interplay between fairness and randomness in a spatial computer game. *International Journal of Computers and Mathematical Learning, 13*(2), 89–110.

Pfannkuch, M., & Ziedins, I. (2014). A modelling perspective on probability. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 101–116). Dordrecht: Springer Science+Business Media.

Piaget, J., & Inhelder, B. (1951). *The origin of the idea of chance in children*. New York: Norton.

Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education, 31*(5), 602–625.

Pratt, D., & Noss, R. (2002). The micro-evolution of mathematical knowledge: The case of randomness. *Journal of the Learning Sciences, 11*(4), 453–488.

Pratt, D., & Noss, R. (2010). Designing for mathematical abstraction. *International Journal of Computers for Mathematical Learning, 15*(2), 81–97.

Prediger, S. (2008). Do you want me to do it with probability or with my normal thinking?—Horizontal and vertical views on the formation of stochastic conceptions. *International Electronic Journal of Mathematics Education, 3*(3), 126–154.

Prodromou, T. (2012). Connecting experimental probability and theoretical probability. *ZDM Mathematics Education, 44*(7), 855–868.

Prodromou, T., & Pratt, D. (2006). The role of causality in the co-ordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal, 5*(2), 69–88.

Prodromou, T., & Pratt, D. (2013). Making sense of stochastic variation and causality in a virtual environment. *Technology, Knowledge and Learning, 18*(3), 121–147.

Ruthven, K., & Hofmann, R. (2013). Chance by design: Devising an introductory probability module for implementation at scale in English early-secondary education. *ZDM Mathematics Education, 45*(3), 409–423.

Schnell, S., & Prediger, S. (2012). From "everything changes" to "for high numbers, it changes just a bit". *ZDM Mathematics Education, 44*(7), 825–840.

Shaughnessy, M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). Reston, VA: National Council of Teachers of Mathematics.

Smith, T. M., & Hjalmarson, M. A. (2013). Eliciting and developing teachers' conceptions of random processes in a probability and statistics course. *Mathematical Thinking and Learning, 15*(1), 58–82.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioural and Brain Sciences, 23*(5), 645–665.

Taleb, N. N. (2010). *The black swan: The impact of the highly improbable*. London: Penguin.

Watson, J., & Callingham, R. (2013). Likelihood and sample size: The understandings of students and their teachers. *Journal of Mathematical Behavior, 32*(3), 660–672.

Watson, A., Jones, K., & Pratt, D. (2013). *Key ideas in teaching mathematics: Research-based guidance for ages 9–19*. Oxford, UK: Oxford University Press.

Weisburg, H. I. (2014). *Willful ignorance: The mismeasure of uncertainty*. Hoboken, NJ: Wiley.

Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal, 5*(2), 10–26.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review, 67*(3), 223–265.

Zahner, D., & Corter, J. E. (2010). The process of probability problem solving: use of external visual representations. *Mathematical Thinking and Learning, 12*(2), 177–204.

Chapter 7
Introducing Children to Modeling Variability

Richard Lehrer and Lyn English

Abstract This chapter synthesizes diverse research investigating the potential of inducting elementary grade children into the statistical practice of modeling variability in light of uncertainty. In doing so, we take a genetic perspective toward the development of knowledge, attempting to locate productive seeds of understandings of variability that can be cultivated during instruction in ways that expand students' grasp of different aspects and sources of variability. To balance the complexity and tractability of this enterprise, we focus on a framework we refer to as data modeling. This framework suggests the inadvisability of piecewise approaches focusing narrowly on, for instance, computation of statistics, in favor of more systematic and cohesive involvement of children in practices of inquiring, visualizing, and measuring variability in service of informal inference. Modeling variability paves the way for children in the upper elementary grades to make informal inferences in light of probability structures. All of these practices can be elaborated and even transformed with new generations of digital technologies.

Keywords Data modeling • Variability • Representations • Informal inferences • Inquiry • Posing questions • Investigating • Attributes • Measures • Data structure • Chance variability • Case-based reasoning • Classifier perspectives • Aggregate perspectives • Digital technologies • Young learners

R. Lehrer (✉)
Department of Teaching and Learning, Peabody College, Vanderbilt University, Nashville, TN, USA
e-mail: rich.lehrer@vanderbilt.edu

L. English
Faculty of Education, Queensland University of Technology, Brisbane, QLD, Australia
e-mail: l.english@qut.edu.au

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_7

7.1 Introduction

This chapter synthesizes diverse research to investigate the potential of initiating elementary grade children into the statistical practice of modeling variability. "Statistics in practice resembles a dialog between models and data" (Cobb & Moore, 1997, p. 810), a contention well supported by social studies of professional statisticians (e.g., Hall, Wright, & Wieckert, 2007; Pfannkuch et al., 2016; Wild, 2006; Wild & Pfannkuch, 1999). Accordingly, we aim to trace potential roots and pathways for bringing children into this dialog. Doing so requires taking a genetic view of the development of knowledge, reflecting a "commitment that the structures, forms, and possibly the content of knowledge is determined in major respects by its developmental history" (diSessa, 1995, p. 23). Moreover, we view developing statistical concepts (such as distribution or statistic) and learning to participate in modeling practices as inherently co-constituted. Hutchins (2012) suggests the construct of *concepts in practice* to emphasize that changes in forms of participation in practice are accompanied by changes in concepts and vice versa. In professions such as statistics, modeling practices are embedded within particular configurations of cognitive, social, and material forms (Knorr Cetina, 1999). Models are invented and contested within a larger system of communal goals, representations, materials, norms, and settings. But these disciplinary aspects of modeling can only be approximated in schooling, and some of them may be so distant from children's experiences that they are poor candidates for instruction. Hence, our examination of research favors seeds of modeling that appear to be essential to the conduct of modeling, that are accessible to young students, and that are potentially capable of growth throughout schooling, if systematically cultivated.

Statistical models are developed as accounts of variability. Yet conceiving of variability is a multifaceted enterprise that includes imagining or participating in the process of creating a sample, visualizing and measuring distribution, differentiating between causal and random sources, and putting variability to use in making predictions and inferences (McClain & Cobb, 2001; Reading & Reid, 2010; Reading & Shaughnessy, 2004; Ridgway, 2015). Studies of professional practice underscore that variability is initiated and interpreted within cycles of inquiry (Wild & Pfannkuch, 1999). For example, in a study of conversations between a statistical consultant and scientist clients, Hall et al. (2007) reported that the consultant foregrounded the program of inquiry established by the scientists, repeatedly diverting their focus on using a particular statistical method to reflecting back on "… again the question you're you're asking… back, always back to that … I mean, what what's the question?" (pp. 110, 113). Wild (1994) also notes that the nature of the question typically provokes reflection about the qualities of a system that are worthy of attention and establishes the need to measure these qualities. Hall, Stevens, and Torralba (2002) describe how a negotiation between a statistician and a team of entomologists resulted in a new measure of differences among chemical profiles of insects. This measure, in turn, transformed qualitative judgments made by entomologists to quantities that could more readily serve to differentiate numbers of distinct colonies within a species, eventually altering the nature of investigation by the entomologists. In related efforts that employed interviews of statisticians about their practices, Pfannkuch and her colleagues (Pfannkuch et al., 2016; Wild & Pfannkuch, 1999) clarified that when conducted by professionals, modeling variability entails coming to grips with multiple senses of variability.

It is clear from these studies of professional practice that if we were to follow statisticians for any significant duration, we would trace a dense configuration of institutional settings, conceptual and material tools, financial arrangements, collaborators, and competitors. We should not ignore these complexities of practice as we consider images of statistical worlds, but because we are concerned with introducing children to these systems and ways of thinking, our goal is to support children's development in ways that are pedagogically fruitful and tractable. To balance complexity and tractability, we find it helpful to focus on a framework that we call "data modeling." This framework suggests how approximations to professional practices of modeling can be coordinated to support children's developing appreciation of variability and uncertainty (English, 2010; Hancock, Kaput, & Goldsmith, 1992; Lehrer & Romberg, 1996). We outline this perspective in the next section.

7.2 Data Modeling

Figure 7.1 outlines a network of concepts in practice that afford entrée to and prospective pathways of development for thinking about variability during the course of schooling. The intention is to create conditions that engage children in developing productive approximations to each node of the network and in learning to coordinate them to describe and account for variability. By approximation, we mean that the form of practice introduced to children preserves its professional function but not its professional complexity. For example, formal treatment of probability density is not typically a target of K-6 instruction, but inference in light of data is, so educators often characterize their efforts as promoting "informal inference" (Makar & Rubin, 2009). Informal inferences go beyond particular cases to make generalizations that recognize uncertainty of the inference, perhaps by using linguistic hedges such as "may" or by referring to a neighborhood of values.

With approximation in mind, the upper portion of Fig. 7.1 addresses practices involved in the conduct of inquiry, ranging from posing researchable questions to deciding what about a system is worthy of measure and designing investigations that will generate a sample of observations. The double arrows depicted in the figure indicate a network of mutually constituted activity and understandings. For example, attempts to characterize attributes of a system guide the development of measures, and the design of measures often makes problematic the nature of the attributes being measured. Questions, attributes, and measures act in concert to inform the design of investigation, including the nature of observation, the selection of units of observation, and the material arrangement of conditions to facilitate observation. These elements are deployed to generate a sample representing the

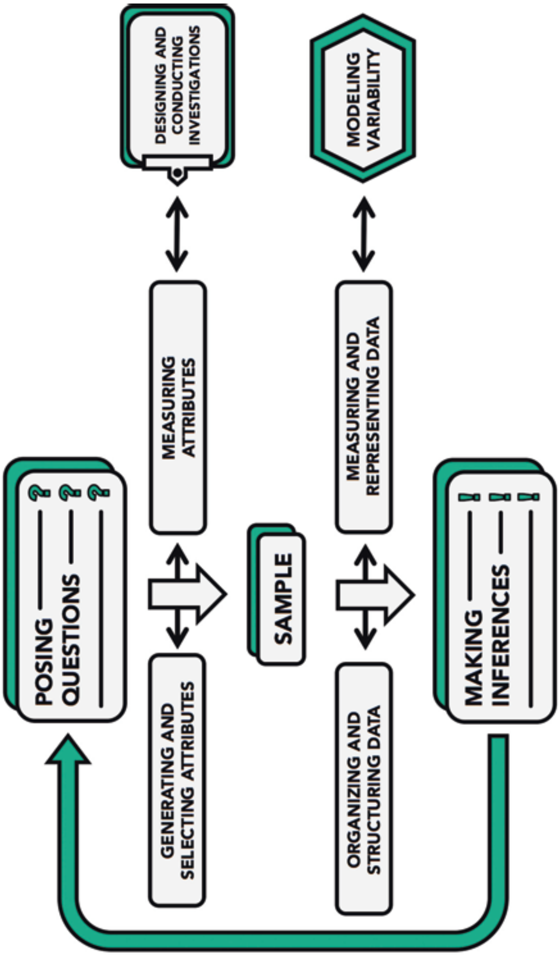


Fig. 7.1. Components of data modeling (adapted from Lehrer & Schauble, 2004)

population defined by the characteristics and measures of the process of interest. It is here that variability is manifested.

The lower portion of Fig. 7.1 refers to structuring, representing, measuring, and modeling the variability now evident in the sample. These forms of activity are also mutually dependent. For example, visual displays of data often suggest that some statistics may be more appropriate measures than others. Models of chance are the principal means of expressing uncertainty about inferences. A sidenote may be helpful here. We are often asked: What is a model? How is it different from a representation or measure? Our perspective is that representations of variability take many forms, including visual display, data structures, measures, and models, and each of these is a way to get a grip on the multiple senses of variability noted previously. Data modeling refers to the assembly of these representations, what Latour (1999) characterizes as circulating reference, to guide inference. Within this network, a *model* is a system composed of random and causal components that stands in for the process that generates the variability observed in the sample (Gould, 2004). Statistical models, unlike other forms of modeling, have a probability structure, ideally one that is made explicit in the model (Pfannkuch et al., 2016).

Note, too, that the capstone of modeling is inference, and in statistical practice, two distinct, albeit often coupled, approaches are apparent in professional practice. The first approach, visually-guided inference, is primarily accomplished by constituting images of trends of variability in a sample. These trends are often rendered narratively to create a causal rhetoric supported by patterns made visible by image (Cobb & Moore, 1997; Kosara & Mackinlay, 2013; Rodgers & Beasley, 2014; Rosling, 2010; Segel & Heer, 2010; Tufte, 1983, 1997). As an example, consider the now-classic image of the distribution of global wealth in the form of an evocative champagne glass, as displayed in Fig. 7.2. This image informs a reader about the global distribution of wealth at a glance, in part through the ironic association provoked by the champagne flute, a vessel for the vintage of the wealthy (Champkin, 2014). The inequality in distribution of wealth is manifest, and concerns about uncertainty, perhaps most pronounced at the boundaries of the regions, seem pointless.

With digital technologies, new opportunities for storytelling arise through dynamic images. For example, with Gapminder an analyst can animate cases and regions of data (Rosling, 2010; Rosling, Ronnlund, & Rosling, 2005). These animated images are often cues to narrative. For example, Hans Rosling[1] recruits narrative devices, such as personification (e.g., relating his family's history to the aggregate), time jumping (e.g., juxtaposing two different periods in time), and sportscast-like metaphors of racing to maintain viewer interest and to clarify complex patterns in global data (Kahn & Hall, 2016). As we later describe, telling stories about visualizations of data provides avenues for inducting children into practices of inference. The second approach to inference relies on explicit probabilistic modeling of the uncertainty that arises from sample-to-sample variability, which in turn requires conjecturing the nature of the stochastic process generating

[1] https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen?language=en)

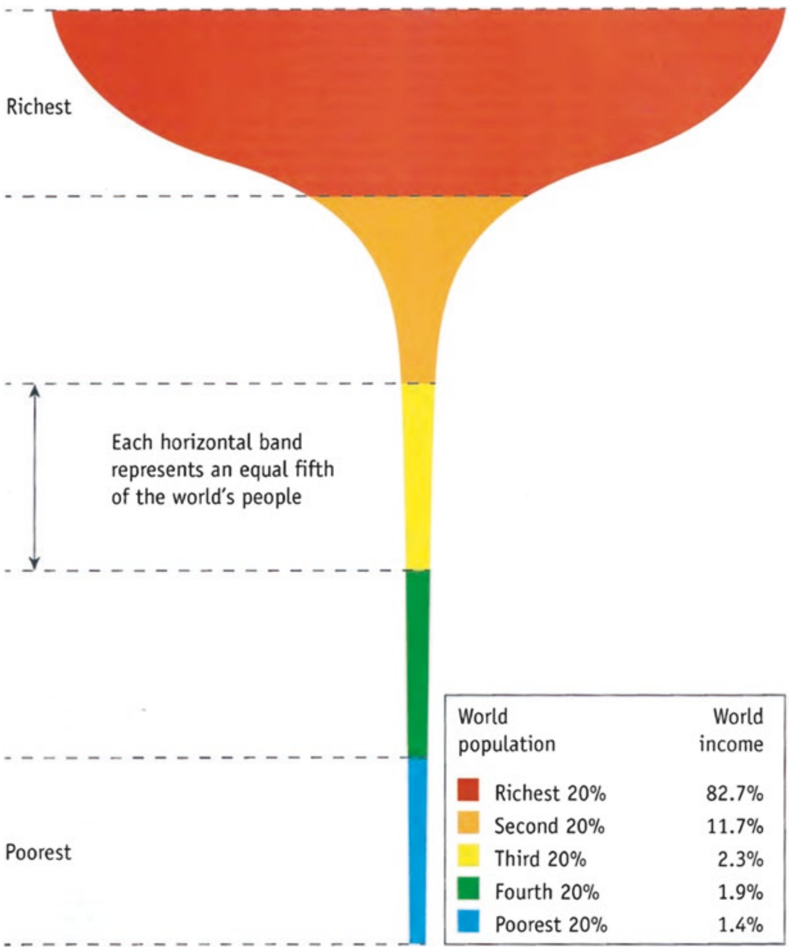


Fig. 7.2. UN report graphic on global wealth distribution (Champkin, 2014)

observed samples (e.g., Pfannkuch et al., 2016; Wild & Pfannkuch, 1999). This face of inference is associated with hypothesis testing, confidence intervals, and making decisions in light of uncertainty. Visually guided and model-based forms of inference are usually coordinated because decisions made in light of stochastic models are themselves anchored in theories and models of the world, so that cause and chance are typically synchronized (e.g., Wild, 2006).

7.3 Introducing Data Modeling to Children

Having introduced this network of aspects of variability, we turn now to examine evidence about the feasibility and intelligibility of introducing children to practices of data modeling. Competence in data modeling is increasingly urgent because young children now routinely encounter a wide range of data in an increasingly diverse space of media. However, in these media, the data construction process is often obscured or is simply unavailable, and claims are often made without considering variability and uncertainty. Schooling in the early years should support children to participate in practices of data modeling so that they are in a better position to appreciate and even participate in this increasingly data-centric world.

One powerful approach to developing these core competencies entails positioning children to participate in multiple aspects of data modeling spanning diverse contexts and associated processes. Consistent with the view of data modeling illustrated by Fig. 7.1, these aspects include (a) posing statistical questions within meaningful contexts that highlight variability; (b) generating, selecting, and measuring attributes that vary in light of the questions posed; (c) collecting firsthand data so that children encounter decisions about the design of investigations; (d) representing, structuring, and interpreting sample and sampling variability; and (e) making informal inferences in light of all these processes. Making informal inferences includes recognizing uncertainty, detecting variation, and making predictions (English & Watson, 2015; Lehrer & Romberg, 1996; Lehrer & Schauble, 2002a, 2002b; Makar, 2016; Watson & English, 2015). In the sections that follow, we examine studies that provide children with opportunities to participate in these aspects of data modeling.

7.3.1 Grounding Data Modeling in Inquiry

7.3.1.1 Posing Questions

A statistical question is the starting point for any investigation; data are generated in a context of inquiry (Moore, 1990). Yet, posing questions is underrepresented in many elementary curricula (Allmond & Makar, 2010; Lavigne & Lajoie, 2007), perhaps because children's initial questions are often informal and broad (English, 2014a; Whitin & Whitin, 2011). Children often find it difficult to generate questions that can be investigated or to envision the data that can address their questions (Allmond & Makar, 2010). Lehrer and Schauble (2002b) suggest that many of the challenges of posing questions for children can be ameliorated when children are given sufficient opportunity to build familiarity with the phenomena being investigated, including opportunities for observation, conversation, and texts that address the target phenomena. As children's interest is cultivated, teachers support inquiry by encouraging children to "collect, categorize, and evaluate the questions posed by the group" (p. x). Teacher support is critical for cultivating a disposition to ask. For example, Allmond and Makar (2010) engaged 9-year-old children in generating and differentiating among questions that could be investigated from those that could not. Children also tried to envision the data that could address each question, and they were encouraged to collaboratively refine questions of description (e.g., "How many peaches are in a can?"), to generate questions of greater collective interest (e.g., "Is there the same amount of peaches inside [every can]?"). After completing several iterations of this process, students were far more likely to generate questions that could be investigated.

Iteration of questions in response to changing conditions of investigation also appears critical. Older children (age 11) investigating aquatic ecologies began with very broad or even irrelevant questions, such as “Who lives here?” or “How much blood can a leech suck?”. As they continued their investigations and became more familiar with characteristics of the ecosystem, their questions began to reflect higher disciplinary importance, such as, “Is the animal life in Pond 2 more diverse than in Pond 1?” Moreover, their questions were more tractable for investigation in light of the material, social and cognitive resources at their disposal. The classroom teacher supported an aesthetic dimension by repeatedly soliciting students’ judgments and justifications about the qualities of “good” research questions. As students proposed and defended their criteria for good questions, the teacher publically posted the criteria that achieved widespread consensus (Lucas, Broderick, Lehrer, & Bohanan, 2005). Students’ criteria shifted over the academic year from an early focus on whether questions were tractable (“Genuine, we don’t already know the answer”; “Doable”) to a growing concern with collective accountability (“People can piggy-back on the question, build on previous questions”) to whether questions supported knowledge sharing across the classroom community (“The answer to the question contributes toward everyone’s understanding”). The trajectory of student questions increasingly reflected productive disciplinary values of collective inquiry.

A further and less studied aspect of children’s question posing concerns their willingness to treat sample data as objects of secondary inquiry. For example, 10-year-old students were surprised to find that the sample data they had generated from a survey of their own design could also be used to address questions that were not originally posed in the survey (Lehrer & Romberg, 1996). Since questions typically arise during conversation, it may be challenging for children to conceive of responses to questions as being subject to inquiry from a source other than a respondent.

7.3.1.2 Developing Attributes and Their Measures

Identifying the attributes that are best for addressing a question of interest necessitates “seeing things in a particular way, as a collection of qualities, rather than intact objects” (Lehrer & Schauble, 2007, p. 154). For example, kindergarten children ordering several pumpkins by “size” had to grapple with what size meant, some proposing height and others, “fatness” (circumference). Having proposed these characteristics, children found that comparisons relied on developing collective understandings about what each characteristic meant. Similarly, first-, second-, and third-grade children who were investigating the growth of organisms decided first about which aspects were the best indicators of growth, such as the “fatness” and length of insect larvae or the height and canopy volume of plants. To support comparisons across organisms, children had to agree about methods and forms of measure, and failure in these methods and forms often led to redefinition of the attributes, as well as the measures (Lehrer & Schauble, 2005; Lehrer, Schauble, Carpenter, & Penner, 2000). Similarly, Manz (2012, 2015) traced how “bumps” (leaf nodes) became accepted in a third-grade classroom as meaningful indicators of plant growth and how production of seeds became agreed-upon indicators of plant success.

In a longitudinal study across grade levels 1–3 (ages 6–8 years), English (e.g., 2013, 2014b) employed storytelling to generate opportunities for children to decide upon the nature of attributes worth attending to by the story characters. For example, as they thought about ways to help a character clean up his room, children developed classifications for items to be gathered and sorted. In the process, they had to agree upon which attributes of the items to attend to, which instigated the need to achieve agreed-upon definitions of attributes. A core goal was providing opportunities for children to focus on the attributes of the items and the ways they varied, so that children could identify, classify, and represent attributes in more than one way. Similarly, diPerna (2002) provided third-grade children with self-portraits drawn by K-5 children. As children considered the variability of the portraits in the set, they generated questions about the differences that they observed in portraits drawn by artists both between and within grades. Of the many questions that were proposed, the children and teacher settled on, “What are some body parts that will show progress [in drawing ability] from pre-K to fifth grade?” (p. 82). To investigate, children were confronted with the problem of generating and defining comparable attributes, just as those in the English (2013) study. This proved challenging, and they initially tended to describe body shapes as “stick figures” or “bubbled out.” These first descriptions proved too global and indistinct, so children refined them to include how the artists drew hands, eyes, and hair. Once students agreed on attributes that seemed to differentiate the portraits drawn by different-aged artists, the practical and critical problem of measure became evident. The categories of eyes or hair or nose that children proposed were as plentiful as the number of portraits. For example, children proposed 14 different eye shapes—an approach that made comparison across grades very difficult. Eventually, children settled on a more tractable set of categories, including “football-like” and “circle-like” eyes. These and other findings suggest that the advice of Hanner, James, and Rohlfing (2002), who replicated this study across grades 1 through 6, is still appropriate:

> Very often, teachers solve all the interesting issues for kids and present them already resolved to children, without giving children the opportunity to grapple with such questions as, “What attributes should we include?” How many attributes should we consider?” and, “How should they be represented?” When teachers take over these decisions, all that’s left is a cut-and-dried graphing or sorting activity, in which teachers have done all the intriguing and motivating thinking ahead of time. (p. 106)

Designing and Conducting Investigations. If students are to overcome the difficulties in linking questions to data (Hancock et al., 1992), questions must contain seeds of investigation that are within the reach of children but not within their immediate grasp (e.g., Allmond, Wells, & Makar, 2010; English, 2010; Lehrer & Schauble, 2002a, 2002b; Makar & Rubin, 2009). Questions must motivate progressive cycles of defining attributes and considering their measure. And, as we noted previously, if the cycle of question posing is sustained for more prolonged periods of time, there may also be opportunities for refining questions in response to the changing status of attributes and their measures (e.g., Lucas et al., 2005). More prolonged periods of time also create opportunities to engage children in the design and conduct of inquiry. For example, fifth-grade students designed a survey to compare their lives with those of early colonists. As they generated and developed these data, students came to understand both the virtues of representative samples and, by looking across multiple classroom responses, the inevitability of sample-to-sample variability—in this case, different proportions of boy and girl respondents across the classes of the school (Lehrer & Romberg, 1996). Cotterman, Lehrer, and Schauble (2014) observed similar change in motivating questions and attributes and measures as sixth graders conducted an investigation of a local creek ecology. Students' experiences with variation in measures instigated practical and conceptual changes in what they considered as a representative sample. At the onset of their investigation, students did not consider inference from a single sample problematic, but by the end of their investigations, they advocated for multiple samples as necessary in light of sampling variability. Similarly, at the onset of investigation, a sample's location was not considered important, yet students soon embraced the need to create representative samples by partitioning the space of the creek to make inference about ecosystem functioning. As they collected data, students debated about whether the absence of an observation should be considered as a valid value in a sample.

Other potential payoffs for longer-term investigation include opportunities for students to come to understand the grounds of experiment, to develop protocols for observation, and to consider how choices of tools and techniques influence what one considers as a sample (Lehrer, Schauble, & Lucas, 2008; Lehrer & Schauble, 2012; Manz, 2012, Manz, 2015; Watson & English, 2015). These findings about sample and sampling variation are not typical in the thinking of elementary school children or even older students, who often prefer a census to a sample (e.g., Jacobs, 1999), conduct biased sampling to ensure the collection of attributes of interest (Schwartz, Goldman, Vye, & Barron, 1998), and generally fail to recognize sampling variability (Rubin, Bruce, & Tenney, 1991). Moreover, many students do not link chance to sampling and fail to appreciate the role of chance in creating representativeness and minimizing sampling variability (Ben-Zvi, Bakker, & Makar, 2015; Schwartz et al., 1998; Watson & Moritz, 2000).

In sum, whether longer or shorter term, investigations that involve students in posing questions, developing attributes and their measures, and generating data firsthand all contribute to articulating data as constructed, not simply as given or as arising from some remote process. Data construction is also a gateway to noticing variability, at first as simple differences in values of measured attributes, even in comparatively sparse contexts such as "how we wake up in the morning" (Putz, 2002), traffic patterns on local streets (Gavin, 2002), planning for a picnic (English, 2011), what canines eat (English, 2013), and personal preferences for varieties of peaches (Allmond et al., 2010). Children also notice values that do not make sense or that are inappropriate, given the context (English, 2012). In the next section, we examine children's conceptions of variability when they have more extensive opportunities to structure it.

7.3.2 *Structuring Variability*

Children's initial perceptions of variability may be restricted to simple indexes of events that generate variability, as in, "We said our favorite colors" (Konold, Higgins, Russell, & Khalil, 2015, p. 309). However, as we noted in the preceding section, noticing different values of a measured attribute, a "case-value" perspective (Konold, Higgins, et al., 2015), is more commonplace. We turn attention in this section to studies that seek to support children in going beyond the case to reason about data as an aggregate. Aggregation is an important stepping-stone toward conceiving of data as distributed (Konold, Higgins, et al., 2015; McClain & Cobb, 2001).

7.3.2.1 Visualizing Variability

In this section, we focus on what children learn by inventing, revising, and contrasting representations of variability (e.g., English, 2013; Lehrer & Schauble, 2002b). One strand of research focuses on inventing displays as tools for promoting representational and meta-representational competencies (diSessa, 2004; Greeno & Hall, 1997). To illustrate, the upper panel of Fig. 7.3 displays a facsimile of a display of silkworm larvae lengths measured at a particular day of growth. The display was invented by third-grade (age 8) children. The lower panel represents all 261 measurements that children generated as a TinkerPlots case-value plot. Children took these measurements as they participated in a unit on the social origins and impact of the commercial production of silk (Lehrer, 2011; Pellegrino, Wilson, Koenig, & Beatty, 2014). Notice that the children's invention emphasizes the value of each individual case and its use of oval icons for each mm of length is a reminder of the morphology of the larvae. However, it also tends to treat space nonuniformly, and so at a glance, lengths of the same measure can appear to have different values (see also, Cengiz & Grant, 2009). There is no such ambiguity in the TinkerPlots display, which also occupies a relatively compact area in contrast to the large portion of the classroom wall that was occupied by the student invention with all 261 measurements represented. Regardless of these advantages of the digital display, the paper technology provided an important pedagogical opportunity for students to develop representational competence about the use of space. As they reviewed the display shown in the figure, about what different inventions made visible (what they "show") and what they tended to reduce ("hide"), several children in the class suggested that the icons needed to be of uniform size.

Figure 7.4 is a facsimile of another display invented in the class, and notice that it, too, uses space nonuniformly. But it also makes use of a classifier (an interval) and a count to create a different shape for the same data. Frequency represents an imposition of structure that is not available in the case-value perspective (Confrey, 2011). This invention makes the center clump of the data more visible, although the center clump can also be seen as a plateau in the case-value plot. As the class

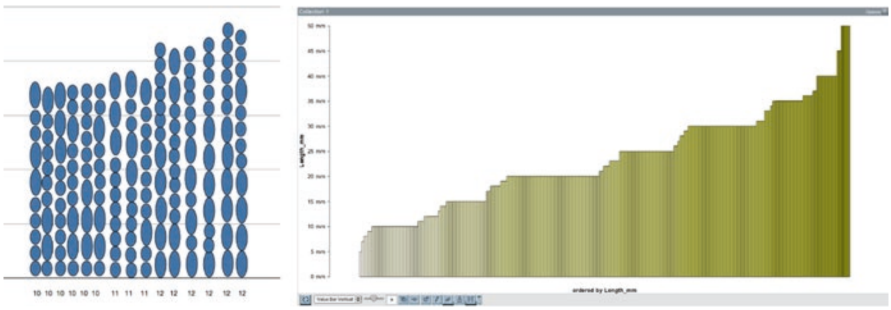


Fig. 7.3 A portion of a case-value visualization of the lengths of 241 silkworm larvae invented by a pair of third-grade (age 8) children to represent variability (*left panel*) and its TinkerPlots counterpart (*right panel*)

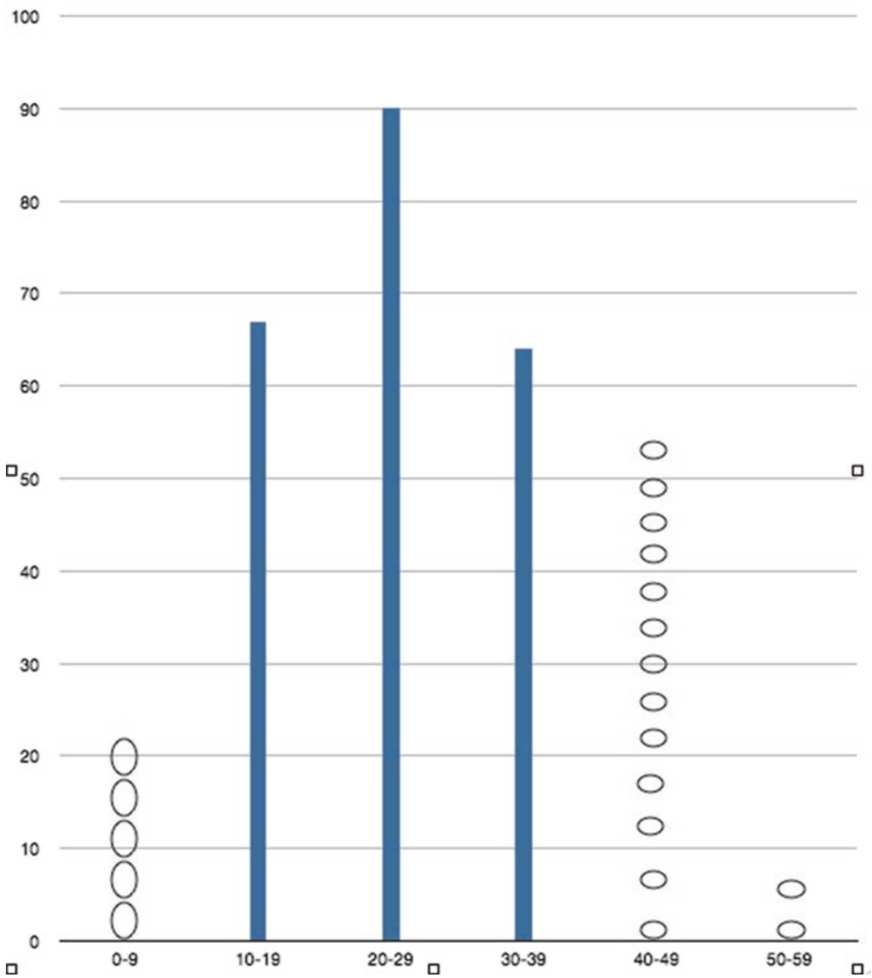


Fig. 7.4. Child invented display of the lengths of 261 silkworm larvae at a particular day of growth

compared these displays, the teacher asked children to take the perspective of the inventors and to conjecture about how the choices made by designers resulted in different shapes for the same data. Following the conversation about showing and hiding, one of the inventors of the case-value display wrote in his journal (invented spellings are replaced here with conventional spellings): “We chose (to make) this because it was easy to understand. It was hard to make all the silkworms evenly … Next time I will make numbers (a reference to frequency).”

Children’s journal entries were often anchored to episodes in the classroom, recalling, for instance, when children compared some of the longer larvae to their teacher’s pinky. These comparisons, along with other invented displays that used the scale of measurement to show “holes” in the data, reflect an aggregate perspective (Konold, Higgins, et al., 2015) in which children attended to the entirety of the data. The discussions prompted several children to envision scenarios of competition in which early-hatching larvae were able to consume more resources than later-hatching larvae. Imagining competition for resources is a valuable conceptual tool for understanding variability within and between species and illustrates the importance of narrative-like interpretations of displays, for both children and professionals alike.

A second, related strand of research suggests that invention and comparison of displays also promotes transitions from case to aggregate perspectives (Bakker & Gravemeijer, 2004; English, 2014a, 2014b; Lehrer & Schauble, 2004). For example, 8-year-old children developed investigations to learn about their classmates’ views of a new school playground (English, 2014a, 2014b). Working in nine different student groups, the children created four survey questions, and the class collectively responded to a subset of these questions. Children were encouraged to represent the class responses in more than one way. Seven created two or more representations and one group created four. The diversity of invented representations reflected the range of perspectives on data noted previously, and this variability helped children understand the value and motivation of perspectives that they did not generate. Similarly, Lehrer and Schauble (2004) challenged fifth-graders to invent displays of plant heights at a particular day of growth, and subsequent class conversations revealed the mathematical procedures inventors employed (e.g., count, order, scale) to generate different shapes for the same data. This variability prompted several students to shift from case to classifier and/or aggregate perspectives. Bakker and Gravemeijer (2004) noted a surge in (seventh-grade) student talk about shape of the data when they were encouraged to invent their own representations of a collection of measures of student weights. Cengiz and Grant (2009) found too that as elementary children compared and contrasted different representations of data, they came to appreciate the role of scale in determining the shape of the data.

Digital technologies, such as TinkerPlots (Konold & Miller, 2011), offer new avenues for visualizing data and supporting the transition from case-based to aggregate conceptions (Bakker & Gravemeijer, 2004; McClain & Cobb, 2001). Supports for distributional (aggregate) thinking include enabling students to organize the data as cases, consistent with their common starting point of viewing variability as simple difference. With TinkerPlots, related tools for structuring data afford re-

representation of cases as collections of similar values, consistent with a classifier perspective, or as positions along a dimension of possible values, consistent with an aggregate perspective (Bakker & Gravemeijer, 2004; Cobb, McClain, & Gravemeijer, 2003; Konold, 2007). Hammerman and Rubin (2004) claim that these features allow learners to get a handle on variability by reducing it into manageable chunks, as in the example of the binning function in TinkerPlots. Binning results in collections of similar values but still maintains the aggregate (e.g., compare the displays in Figs. 7.3 and 7.4).

In summary, positioning children to participate in cycles of representational redescription of data collected for some meaningful purpose encourages development of multiple perspectives on data. These multiple perspectives often emerge as children consider how the mathematics of count, interval, and scale can be put to use to highlight what they have noticed. As children invent, they have opportunities to consider how the shape of the data, its visual appearance, is influenced by use of these mathematical means. For example, "center clumps" and related guideposts to interpreting variability are much less visible in case-value displays than in dot plots. It is important that visualizing data informs children about something that they do not already know about the data. The discovery of something new is one of the primary motivations for visualizing data in professional practice. However, as Konold, Higgins, et al. (2015) caution, "… the perspectives one takes on data should serve one's questions rather than the other way around" (p. 323). Hence, although aggregate and classifier perspectives have dominated our presentation of children's efforts to represent, it is useful to recall that case-value perspectives are not always mere starting points to more sophisticated ways of representing. Children often view case-value displays as persuasive ways of representing variability, for example, looking for "plateaus" in case-value plots as indexes of low variability in regions of data (e.g., Lehrer & Schauble, 2004). In other situations, as well, case values may be better representations in light of the question. After all, meta-representational competence means having a wide repertoire of perspectives and selecting appropriately among them (diSessa, 2004).

7.3.2.2 Structuring Variability by Measuring

Children typically treat statistics as computational artifacts, rather than as measures of characteristics of distribution (e.g., Bakker, 2004; Watson, 2006). Watson (2006) mentions that the mean is often privileged, so that other measures of central tendency, such as median and mode, are "often portrayed as poor relations of the mean" (p. 121). Yet in contexts of data modeling, children can come to understand statistics as measures of distribution that guide inference about questions (Makar, 2014). One way to support this conception is to position children to invent statistical measures of center and spread and to evaluate these measures as indicators of a distribution's characteristics (Bakker & Gravemeijer, 2004; Konold & Pollatsek, 2002; Petrosino, Lehrer, & Schauble, 2003). Konold and Lehrer (2008) recommend that repeated measure and production processes characterized by signal and noise support making sense of statistics as measures. In contexts like these, measures of center correspond to signal, and measures of variability correspond to noise (English & Watson, 2015; Konold & Pollatsek, 2002; Petrosino et al., 2003). Signal and noise in these contexts can both be considered as caused, even as variation about them can be considered as arising due to chance, beyond the control of individual agents. This duality is an important resource for learning to think statistically (Biehler, 1999).

For example, Prodromou and Pratt (2006) report how two students interacted with a microworld designed to promote a link between cause and chance. The microworld consisted of random generating devices, such as spinners and dice, which were described as defective, so that students could fix them in contexts involving signal and noise (e.g., a basketball player trying to make shots). Their approach was to provide digital tools so that students could influence the probabilities of events. As one pair of students (14 years old) noted: "When the arrows (sliders in the microworld) are close together, it's got more a chance of going into the net" (p. 83, parenthesis added). This remark reflects a coordination between cause and chance.

In Petrosino et al. (2003), fourth-grade children invented a measure of variability, a "spread number," to describe the height of the apogee of a rocket as determined by multiple measurers. They did so by finding the median of the absolute values of differences between each measured apogee and the sample median. Students noted that a value of zero would correspond to all measurers getting exactly the same value, a comment that indicates they are regarding the statistic as a measure. They also noticed that as their methods of measurement improved, the spread number decreased. These observations provided an avenue for coordinating cause (change in method) with chance (residual error variability). Similarly, fifth- and sixth-grade students measured the arm span of their teacher with two different tools, a 15 cm ruler and a meter stick. They noticed that the change in tools caused a change in the variability of the resulting collection of measurements. Nonetheless, no matter how "carefully" students measured and how assiduously they avoided "mistakes," they found although they could influence the magnitude of variability, they could not eliminate it (Lehrer, Kim, & Schauble, 2007). Konold and Harradine (2014) suggest that, in contexts of repeated measure and production, "… because we are in control of these processes, we can minimize variability" (p. 242). Moreover, in these contexts, students can provide "detailed descriptions of process components that produce variability" (p. 240).

Whole-class critiques of student-invented measures can play an essential role by affording opportunities for students to analyze the grounds of proposed measures and to consider generalization to distributions that can be conceived, even if they were not explicitly generated during the conduct of investigation (Lehrer & Kim, 2009; Lehrer, Kim, & Jones, 2011). Conceiving of sample distributions consistent with a process but not yet realized (in real data) is likely an important resource for developing understanding of sampling variability (Thompson, Liu, & Saldanha, 2007). Another virtue of signal-noise contexts revealed by these studies is that the processes generating them are accessible to students, so that they can anticipate why, for example, a change in a method of measure might affect the value of a measure of variability much more substantially than it will a value of a measure of center (Lehrer et al., 2007).

In summary, traditional instruction often treats measures of distribution, such as characterizing its center and variation, as matters of computation. But there is clear evidence that children can come to recognize statistics as ways of measuring characteristics of distribution. Bakker and Gravemeijer (2004) claim that when students do so, they are viewing data through the lens of distribution, in contrast to a mere set of data values. It may be that contexts of signal and noise have unique affordances for introducing children to statistics as measures, although other contexts involving meaningful (for children) social processes and even natural variation may also have advantages (Ben-Zvi et al., 2015).

7.3.2.3 Organizing Variability as Data Structure

Most research focused on introducing students to data modeling either tacitly or explicitly employs a case-by-attribute table structure. Hancock et al. (1992) noted that children often find this format challenging and prefer to associate individual cases with separable sets of values. For example, as they recorded data about the gender and names of a group of people, children preferred to sort the names into the two genders rather than to code them using the attributes *name* and *gender*. In a study of student-generated data models to predict the age-grade of artists who drew pictures of houses near and far, young children in the study (grades 1, 2) generated unique attributes and values at each grade level. In contrast, older children (grades 4, 5) were able to employ common attributes with multiple values that spanned the range of the artists' grades. This dimensional structure allowed older, but not younger, children to create predictive models based on combinations of attributes and values that spanned the ages of the artists (Lehrer & Schauble, 2000). The older children's creation of a case enabling comparison across units of observation is "at the very foundation of data modeling" (Konold, Finzer, & Kreetong, 2015, p. 4). In the Konold et al. (2015) study, participants at the middle and high school level, as well as adults, reviewed schematics of traffic flow with variables such as time and date and vehicle type. Then they created a data organization that would assist city planners. The youngest members of the sample spontaneously, and in a brief span of time, created narratives that bound information on multiple variables together into cases, as in "Car 4, going at 50 mph, 30 feet behind car 3." Thirty-seven percent of these students created tables, most of them reflecting a nested structure, not the flat, case-by-attribute structure that is most prevalent in applications and in education. Hence, it appears that most forms of data structure employed in elementary schools do not align well with how children or even older students tend to think about the organization of data.

7.3.2.4 Structuring Chance Variability

One of the primary contributions of statistics consists of ways and means of structuring random variability. Yet children and even older students usually consider variability deterministically (e.g., Ben-Zvi, Aridor, Makar, & Bakker, 2012; Metz, 1998). For most children, chance is a synonym for haphazard, and they often believe that personal agency (e.g., lucky numbers) can explain any structure in variable events (e.g., Horvath & Lehrer, 1998). Many also interpret probability as a description of a single outcome (Konold, 1989). Saldanha and Thompson (2002) indicate that a conception of repeated random process is foundational for developing conceptions of probability, and several studies explore children's conceptions of probability when they engage in exploring and explaining the behavior of simple random processes.

Fielding-Wells and Makar (2015; see the description in Chap. 8) focused on shifting young children's (7–8-year-olds) notions about equiprobability, that is, believing that all outcomes of a chance process must be equally likely (Hawkins & Kapadia, 1984). Children played Addition Bingo, which involved generating all possible combinations of the sum of two numbers (1–10), each written on a slip of paper and placed in a box. Children were given a card consisting of a 5 × 5 array of self-selected numbers, that is, their predictions of which number sums would be called, allowing for repeated numbers. As each sum was drawn (e.g., 2 + 9) from the box, the children marked off the sum (in this case, 11) if it appeared on their card. The player who was the first to mark off all of the numbers on his/her card was the winner. While trying to provide evidence of the "best" card that would win the bingo game, the children encountered conflicts between what they expected and the outcomes of the game. Considerations of sample space arose when students noted that the frequencies of each sum they recorded differed from those of their peers. Considering the sample space helped explain this observation. Continued play with smaller, faster versions of the bingo game, accompanied by a dot plot to keep track of the successive numbers called, helped children to envision and anticipate structure in random variability.

This anticipation of structure was also evident in research conducted by Horvath and Lehrer (1998), in which second (age 7)- and fourth-grade children (age 9) first observed and recorded the outcomes of repeated throws of a single six-sided die. Initially, children interpreted differences in the frequencies of outcomes of the die as evidence of their own agency, in line with early understandings of chance variability. For example, a child claimed that higher frequency meant, "Well, it has always been my lucky number, you know" (p. 139). Agency-based reasoning may be derived from a feeling of physical control of the die (Pratt, 2000). As children aggregated results of repeated trials, a growing sample approach (Bakker, 2004), lucky numbers, and related forms of agency began to seem less tenable.

Most children shifted toward expecting equally likely outcomes and treating differences among outcomes as noise instead of as signal. When they began throwing two eight-sided dice and finding the sum, children noticed a new shape emerging. They dubbed it the "mountain" and predicted a similar shape for throws of a six-sided die. A minority of the youngest children related the mountain to "ways of making them" (permutations or combinations), comparing ways of generating sums of 16 and of 9 for repeated throws of two eight-sided dice. Enumerating ways of making was more commonplace and accessible with older, fourth-grade (9 years of age) students conducting similar investigations. Older children were much more likely to explain empirical outcomes by recourse to the sample space. However, at both grade levels, contests between permutation and combination as counting methods for sample spaces were not resolved, perhaps because both resulted in global mountain shapes. Moreover, aggregating across throws of dice provoked considerable debate among the younger, grade 2 children. Some advocated that they should be allowed to throw the die in ways that allowed them to bias the outcomes toward lucky numbers, just so long as the dice did not produce the lucky number all of the time. This was a hybrid of their original sense of control over the dice with their growing realization, emerging from repeated trials as the sample grew, that they lacked control over particular outcomes. Other second-grade children saw this proposal as unfair. They argued instead for a norm that would govern throws that could be combined in ways that would not bias the outcomes. Apparently it is worthwhile to explore circumstances under which one event can be construed as enough like another event to warrant status as repeated. Children's concept of a trial is important to consider in their explorations of probability.

Pratt (2000) claimed that digital microworlds offer opportunities to extend children's experiences of probability and thus provide new resources for learning. Accordingly, he designed a microworld in which students could amend a "working box" that rendered a version of the sample space so that they could construct a digital device that mimicked observed sums with material dice. This form of digital support was supplemented by his interactions with students to support systematic enumeration and to relate enumeration of outcomes to the relative proportion of sums displayed in a pie chart. Pratt suggests that children's initial belief that fairness means the same likeliness of each sum was reconstructed so that fairness came to mean the equal representation of each possible outcome. This emphasis on fairness reconciled variation with expected values in the sums.

Similarly, Abrahamson (2012a, 2012b) described attempts to affect what he called a synthesis between perceptual and disciplined anticipations about the operation of chance. Students in grades 4–6 predicted the outcomes of drawing four balls from an urn that contained equal numbers of green and blue marbles. The four possible outcomes were structured perceptually by a spoonlike utensil with four slots, each of which held either a green marble or a blue marble. Students did not run the experiment, but tended to predict that the most likely outcomes were those involving (two green, two blue). Students then found the possible combinations ([0b, 4 g] [1b, 3 g] [2b 2 g] [3b 1 g] [4b, 0 g] and inscribed each on a card. Abrahamson described interactions with one sixth-grade student, who, after generating the combinations, changed his prediction to equal probability of all outcomes, in accord with this representation. The interviewer-teacher then induced the student to create all possible permutations and to represent them graphically as a "tower" (pictograph).

The representation made it evident that the permutations involving [2b 2 g] outnumbered those of any other combination. Interviewer assistance highlighted the implications of each representation (combinations, permutations) for the prediction of outcomes, and the student eventually concluded that the representation of permutations was consistent with his first intuition. Abrahamson (2012b) suggests that designing so that students can reconcile their intuitions about chance with a disciplined conception of sample space provides an alternative to approaches that rely on conflict between anticipations and empirical results.

English and Watson (2016) also focused on the foundational concepts of variation and expectation as they attempted to support grade 4 (9 years of age) students' understanding of probability. They investigated students' initial expectations of the outcomes of tossing one and two coins, how their expectations changed as they repeated the numbers of trials, and how their use of computer simulation (via TinkerPlots, Konold & Miller, 2011) led them to understand the relationship between experimental estimates of probabilities and theoretical probabilities as the number of tosses increased. Three phases were identified in the development of the fourth-graders' understanding of probability. Initially, students displayed a basic awareness of the uncertainty of chance events and some intuitive appreciation of probability with respect to independence of outcomes and coin type. For example, after tossing a coin once, children predicted the outcomes of another ten tosses. Most did not expect identical outcomes for each toss; approximately 35% were uncertain about the number of heads and tails yet expected about five of each due to "equal chance" (p. 41). When they predicted the outcomes of tossing two coins, the majority of the students' responses aligned with the common equiprobability response of three equal outcomes, each with a probability of 1/3 (Hawkins & Kapadia, 1984). Some students (23%) predicted four possible outcomes, but the majority of these students did not associate these outcomes with probabilities.

During the second phase of instruction, students engaged in structuring and representing the outcomes observed from tossing one coin ten times. Although most students readily observed the shape of the data (e.g., "mountain"), under half of them (45%) associated the central tendency of the outcomes with anticipations of a neighborhood of values about five heads. Limited reference was also made to chance as a reason for the center clump. When students used TinkerPlots to grow the number of trials for tosses of one coin, they were more likely to relate center clump to chance. When students progressed to tossing two coins 12 times, displaying their data from their group experiments disrupted their anticipations of equiprobabilities. It was clear from their experiments that "one of each" ([h, t] or [t, h]) occurred about twice as often as either [h, h] or [t, t.]

The final phase of instruction, formal model construction, involved students in displaying their experimental understanding symbolically and diagrammatically. The actual model construction, however, is only part of the process, as English and Watson stressed. In addition to constructing a model, students must be able to interpret their model, explain what it conveys, and relate it back to their initial investigative question. The majority of children (55%) did not relate probabilities of joint outcomes to the sample space in their models, but a substantial number (44%) displayed the four repeated outcomes as HH, HT, TH, and TT and linked these appropriately to chance combinations of the two coins or to the probabilities of the corresponding outcomes. In the most advanced models, there was evidence of a sophisticated understanding of the probabilities. For example, Dagmar's model listed eight possible outcomes and linked these to the probabilities of 1/4, 1/4, and 1/2. She explained:

> Well, there are two heads. One of the heads is in coin 1 and the other one is in coin 2. Then there are two tails, one of them is in coin 1 and the other one is in coin 2. Then there is one head and one tail, so um, the heads is in coin 1 and the tails are in coin 2 and then there's the same one except they are rotated. The heads are in coin 2 and the tails are in coin 1. (Researcher [R]: And the fractions?] And then the fractions is, the first one, two heads is 1/4, the second one is 1/4 as well, two tails, and then both of them, the last one equals 1/2. [R: How did you work it out?] Well, I said, um … that's 1/4 and that's 1/4, and then I, um, that's an equivalent fraction to 1/2 so yeah, so I just did 1/2 (English & Watson, 2016, p. 52).

In summary, research suggests that when it is carefully designed and supported by teachers, investigating the behavior of simple random devices can help children form an image of a long-term stochastic process and explain the outcomes of these processes with sample spaces, when these are relatively easy to construct and enumerate. There is a marked tendency for students to conceive of sample spaces as combinations, not permutations. Because the studies of children's understandings of chance in these contexts span multiple decades and even continents, yet converge on similar findings, it would clearly be fruitful to engage children in considerations of chance more often and more systematically during the elementary years. The design of such instruction can now include new digital means for representing and experiencing the random variation of simple stochastic processes.

7.3.3 Informal Inference

The cycle of statistical inquiry depicted in Fig. 7.1 terminates and reinitiates with informal inference. For several decades, research has focused on informal inference as characterized by recognizing uncertainty, detecting variation, and making predictions (English & Watson, 2015; Lehrer & Romberg, 1996; Lehrer & Schauble, 2002a, 2002b; Makar, 2016; Watson & English, 2015). In much of the research reviewed to this point, children's efforts to structure variability are grounded in the need to warrant inferences about questions and claims related to their inquiry. Many of the inferences that children make in these contexts rely on noticing characteristics of displays and employing these to justify narrative accounts of, for instance, plant growth or of recycling in a community or of the behavior of a character in a story. Children are also apt to use cut-points and even differences between statistics in service of inferences that fit the criteria for informal inference suggested by Makar and Rubin (2009): these inferences are grounded in data as evidence, go beyond particulars to reason about trends and related generalizations, and typically include recognition of uncertainty.

Makar, Bakker, and Ben-Zvi (2011) further suggest that contrasts between belief and data often amplify children's recognition of uncertainty. In their study of grade 6 students contesting whether sixth- or seventh-grade classmates could jump further, outcomes in favor of sixth graders contradicted prior beliefs and prompted increased efforts to examine the nature of the samples more carefully. For example, students collected more than one sample because one of them noticed that random samples did not guarantee representativeness of gender (Makar et al., 2011). Similarly, fourth-grade students (9 years) in Petrosino et al. (2003) compared the apogees of rockets with rounded and pointed nose cones. They firmly anticipated that the pointed nose cones would cut through the air and travel higher. Dividing a dot plot of values of height at apogee for a rounded nose cone into three "superbins" defined by the median and lower and upper bounds of a "spread number" (a statistic of variability), they found that 51% of the apogee measurements of rounded nose cones fell into this middle bin, and roughly equal percentages fell into the lower and upper bins. Treating the distribution of rounded nose cones as a reference distribution, their teacher invited predictions about where values from the pointed nose cone would fall. Students noted that they could not be certain, but predicted that most values would fall in the upper third of the reference distribution. They were surprised to find that 86% of the values occupied the lowest bin and reluctantly concluded that rounded nose cone rockets tended to go higher. As might be expected from the interplay between cause and uncertainty noted earlier, students had difficulty reconciling this inference with their causal models of air resistance: "I don't disagree, but I think it's kind of weird that the, um, the pointed doesn't go as high as the rounded. That doesn't really make sense" (p. 145).

In other inquiry settings, US fifth-grade students conducted comparative investigations of methods for remembering state capitols. After developing a pretest measure, they compared the effects of a process involving repeated rehearsal ("Sean's method") to the method of loci, a mnemonic method developed by an educational psychologist. Students' initial analyses focused on the number of cases in each condition that exceeded a cut-point, but other students noticed that the rehearsal method produced less variability than did the mnemonic method. This focus on relative variability, coupled with differences in the sample sizes associated with each condition, led students to consider relative proportions at dual cut-points representing fewer capitals recalled and more capitals recalled. This was a form of reasoning aligned with a classifier view of the distribution. It led to an informal inference in favor of Sean, even in light of the uncertainty produced by the variability of outcomes (Lehrer & Schauble, 2002a). These studies suggest that the seeds of statistical inference are within children's grasp in contexts that provide opportunities for inquiry, construction, visualization, and measure of variability, all considered in service of warranting a claim or answering a question.

However, most of the research tends to treat investigation of chance variability and of data variability distinctly. As we noted, many studies of children's conceptions of probability focus on efforts to help children develop sample spaces as explanations of the structure of outcomes for simple random devices. But these devices are rarely positioned as standing in for processes other than themselves. In contrast, integrated modeling environments feature children (usually at upper elementary grades) employing models of chance in service of inquiry about phenomena in which chance variability is thought to be at play (Manor Braham & Ben-Zvi, 2015). We reserve the term modeling for this enterprise—one in which children's understandings of chance and the behavior of chance devices are harnessed to explain observed variability of something other than a random device. That is, the probability structure of the chance device serves as an analogical source for the target domain.

For example, fifth-grade students investigating claims about extrasensory perception designed an experiment in which each student played two roles. One was to throw a six-sided die out of view of a partner while simultaneously concentrating on forming an image of the outcome. The other partner guessed the outcome. This process was repeated 20 times in each of four different conditions, ranging from extreme quiet to extreme noise. As students expected, the mean number of correct matches was highest in the quiet condition, a finding that confirmed their anticipation of the operation of ESP. Their teacher suggested that they model what might happen if the outcome had happened just by chance, a suggestion posed as an alternative explanation for their findings. Students first formed a composite variable for each of the 21 participants to represent the "total effects of ESP" across the 80 trials per participant. Then, they threw the die, recorded the outcome, and repeated the process 79 more times for each simulated participant. They were surprised to find that the distributions of the chance model and the data could not be distinguished and reluctantly concluded that "our high scores could be entirely due to chance" (Lehrer & Romberg, 1996, p. 101). Like good scientists, they were reluctant to give up and suggested refining their hypothesis, so that future work would focus on the outliers in the data! In the next section, we review contemporary efforts to help upper elementary (grades 5–7) integrate models of chance with data to generate informal inferences.

7.3.4 *Model-Based Informal Inference*

Manor Braham and Ben-Zvi (2015) advocate that youth interest should be the centerpiece for modeling. In their study, seventh-grade (age 13) students employed questionnaires to generate data about peers' musical preferences and explored relations among these preferences and other variables, such as gender. To integrate data construction and exploration with modeling, students used TinkerPlots to resample the large sample of questionnaire data to explore the relation between sample size and the confidence they might have in particular sample statistics, such as the mean percent of students who preferred rock-and-roll music. Student-generated models were employed to generate sampling distributions of preference-related statistics (e.g., the value of the sample mean percent preference for rock music from sample to sample), with an eye toward exploring the variability resulting from the same model with varying sample sizes. In a case study of two students, one reasoned probabilistically about an acceptable range of variability in a statistic's value based on its sampling distribution, while the other tended to acknowledge sampling variability, but focused instead on the differences between sample and sampling statistics without quantifying chance.

Konold and Kazak (2008) introduced students to TinkerPlots models as ways of extending their inquiry about whether games—which students first played physically—were fair. Posting the outcomes of different pairs of students as they played each game manually highlighted sample-to-sample variability. Modeling helped students appreciate this variability as a result of a signal of sample space and noise of chance departures from the signal. In further exploration of modeling signal and noise, students (ages 13 and 14) participated in a simulated manufacturing process during a weeklong teaching experiment (Konold & Harradine, 2014). They manufactured "fruit sausages" composed of Play-Doh with two methods of production. To compare methods of production, the researchers introduced a TinkerPlots model of signal and random variation from target as a way of summarizing their previous experiences. Students critiqued the model and compared their expectations of trends that might arise from factors such as practice, which might lead to improvement, with the random error of the model as it generated fruit sausages. The student critique included concerns that the model failed to include sources of error, such as pressing too hard on an extruder during the production process. Nonetheless, modeling provided students with ways of interpreting sample variability and distinguishing between trends in data anticipated from their experiences and the data generated by the model.

In a related context involving signal and noise, fifth- and sixth-grade students invented and revised models of the variability in a sample of repeated measures (Lehrer et al., 2007, Lehrer, Kim, Ayers, & Wilson, 2014). Students understood that the sample observations (the measures they collected) were composed of true measure and random error. In accord with this understanding, students constructed chance devices to represent the magnitudes and likelihoods of different sources of error they had identified, such as the propensity of measurers to leave small gaps when translating a ruler to measure a distance. Students combined the outputs of these chance devices with a constant estimate of the true measure (usually the sample median), as displayed in Fig. 7.5, and ran the model to generate simulated samples and sampling distributions of measures of center (e.g., medians) and of variability (e.g., IQR's). Sampling distributions informed model fitting, and after judging a model as fit, students used the sampling distributions of model statistics to make inferences about the effects of changes in measurement processes.

Interviews were conducted at the conclusion of instruction with 12 students (Lehrer, 2015) and focused on sample statistics arising from claims about either different targets of measurement or improvements to a measurement process. Most students' inferences were guided by reasoning about the probability of the sample statistic in light of the model-based sampling distribution of that statistic. Moreover, most students explicitly recognized that their decision still was uncertain—there

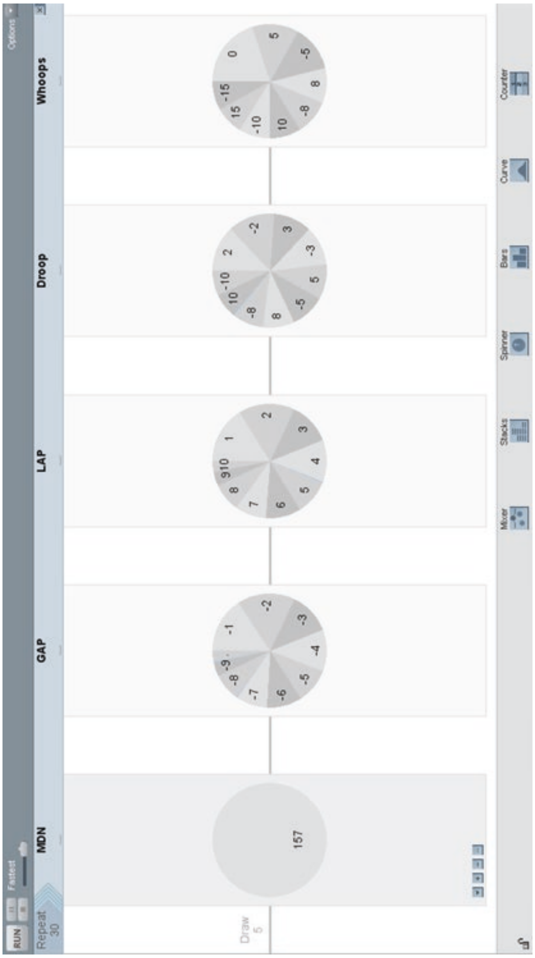


Fig. 7.5. Student signal-noise model of measurement conducted in their classroom

was some chance that a sample statistic could be due to chance, no matter how improbable that might seem, so claims about change in process were always somewhat uncertain. In this investigation, and in others with a similar focus (e.g., Manor Braham & Ben-Zvi, 2015), modeling variability paved the way for students to reason about sample-to-sample variability and, by constructing sampling distributions, to consider claims amid uncertainty.

7.4 Some Directions for the Future

Considered collectively, the studies we examined establish that children are capable of participating productively in forms of modeling data. Modeling data positions students to develop multiple senses of variability and multiple means for getting a grip on it, in the form of visualizing, measuring, and even modeling its production. Even young children profit from constructing data generated in response to a legitimate question, an activity that encourages children to consider and participate in developing relations among question, attribute, measure, and design of investigation. Some instructional settings focus on only a subset of these processes of data construction, but, nonetheless, children's participation in even portions of these processes appears to provide firsthand experience of sample variation, which is a necessary ground of statistical inquiry.

Yet anticipating variability is merely preamble to considering its structure. Studies that engaged children in producing and critiquing visualizations (representations) of variability helped them make informal inferences that were tempered by variability. These findings are especially promising in light of well-documented challenges that both children and older students have in considering samples and variability. For example, children often prefer a census to a sample (e.g., Jacobs, 1999), conduct biased sampling to ensure the collection of attributes of interest (Schwartz et al., 1998), and generally fail to recognize sampling variability (Rubin et al., 1991). To address these misunderstandings and to better align instruction with professional practice, it is desirable that children produce and/or use visualization technologies to produce representations that reveal an aspect of the process investigated that was not immediately apparent to children. It is also important that children come to understand that the qualities of the visualization are a product of choices made by designers. Similarly, engaging children in the invention and critique of measures of variability attunes them to relations between characteristics of a distribution and the measure of those qualities. Statistics may provide opportunities for children to see that measures are not confined to spatial magnitudes, as is the case in most elementary education.

An unresolved issue raised in our review of literature focuses on cultivating images of chance based on repeated processes. This is a unique contribution of statistical modeling, and three different approaches are evident in the studies we reviewed. One approach seeks to immerse children in the investigation of simple repeated processes that are familiar from textbooks on probability, such as repeated throws of a die or dice, games of chance, and the like. These have the affordances of ready participation by children, surprising (to children) outcomes that suggest that chance may have structure, and relatively simple sample spaces. Digital technologies are providing new opportunities and perhaps more effective ways on capitalizing on students' intuitions in contexts like these. A second approach seeks to seamlessly integrate chance with data construction, so that chance is introduced as a way of explaining the variation that children experience firsthand as they construct data. A third approach is a kind of hybrid. It entails engaging children in the exploration of simple random processes and then framing these processes as models of phenomena that children originally may not consider as involving chance. Each approach provides opportunities for children to develop representations and measures of outcomes, but more research is needed to explore the trade-offs that each entails. From a professional practice point of view, integration is preferred, but in keeping with the genetic spirit of our endeavor, integration may or may not be the best way to support children's development.

The emerging focus on probabilistic modeling of variability in outcomes based on popular interests or on their involvement in tangible processes, such as making things with the intention of achieving consistent products, appears to offer children in the upper elementary grades a way to bridge from sample variability to sampling variability. The latter is a key to statistical inference, and so it is encouraging that preliminary explorations suggest some evidence of what Manor Braham and Ben-Zvi (2015) term probabilistic thinking—that is, students are able to use sampling distributions generated as they invent and revise models to frame the uncertainty involved in making even simple decisions. This is especially surprising in light of consistent evidence that many students do not link chance to sampling (Ben-Zvi et al., 2015; Schwartz et al., 1998; Watson & Moritz, 2000). But much more research is needed to better understand how children's, and for that matter, older students' interpretations and uses of probabilistic models are influenced by particular teaching practices and by variations in phenomena being modeled.

In an era of big data and planetary scope, it nonetheless seems critical that children's use of data reflect their fields of inquiry, which are apt to be local. Considering local questions provides an opportunity to cultivate dispositions and values of data modeling. More developmental, long-term investigation of these issues is merited, especially productive pathways that lead to systematic expansion of the scope of inquiry to focus on socially relevant and scientifically consequential issues. It is an unfortunate reality that most instruction about data modeling occurs in the context of mathematics classrooms. Yet much of the power of data modeling is evident in investigations of the natural and social worlds. Indeed, much of data modeling originally arose to address challenges posed by variable outcomes in these contexts (Porter, 1986). The discipline-specific partitioning of children's education works against this historic development. The inclusion of tools and ways of thinking about "big data" and about more complex systems, such as those involving covariation (e.g., Cobb et al., 2003; Chap. 4 this volume), and/or public participation in a variety of data modeling enterprises, such as election polls or state sponsored lotteries (e.g., Rubel, Lim, Hall-Wieckert, & Sullivan, 2016), are all promising pathways for extending the largely univariate and more local scope of investigation typical of data modeling in the elementary grades. Because data modeling is of long-term value to participants in an ever-expanding social and material world, early entry and sustained cultivation will be consequential.

References

Abrahamson, D. (2012a). Rethinking intensive quantities via guided mediated abduction. *Journal of the Learning Sciences, 21*(4), 626–649.

Abrahamson, D. (2012b). Seeing chance: Perceptual reasoning as an epistemic resource for grounding compound event spaces. In R. Biehler & D. Pratt (Eds.), Probability in reasoning about data and risk [Special issue]. *ZDM–Mathematics Education*, *44*(7), 869–881.

Allmond, S., & Makar, K. (2010). Developing primary students' ability to pose questions in statistical investigations. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics. Ljubljana Slovenia, 11–16 July 2010*. Voorburg, The Netherlands: International Statistical Institute.

Allmond, S., Wells, J., & Makar, K. (2010). Peach ponderings. *Thinking through mathematics: Engaging students with inquiry-based learning* (Book 2: ages 8–10, pp. 74–82). Melbourne: Curriculum Press.

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64–83.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Netherlands: Springer.

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM—The International Journal on Mathematics Education, 44*(7), 913–925.

Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics, 88*(3), 291–303.

Biehler, R. (1999). Learning to think statistically and to cope with variation. *International Statistical Review, 67*(3), 259–262.

Cengiz, N., & Grant, T. J. (2009). Children generate their own representations. *Teaching Children Mathematics, 15*(7), 438–444.

Champkin, J. (2014). The champagne glass effect. *Significance, 11*(1), 39–41.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*(1), 1–78.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 103*(9), 801–823.

Confrey, J. (2011). Making sense of practice in mathematics: Models, theories and disciplines. In T. Koschmann (Ed.), *Theories of learning and studies of instructional practice* (pp. 323–336). New York: Springer.

Cotterman, S., Lehrer, R., & Schauble, L. (2014, July). *Sampling in the wild.* Paper presented at the 9th International Conference on Teaching Statistics, Flagstaff, AZ.

diPerna, E. (2002). Data models of ourselves: Body self-portrait project. In R. Lehrer & L. Schauble (Eds.), *Investigating real data in the classroom* (pp. 81–97). New York: Teachers College Columbia University.

diSessa, A. (1995). Epistemology and systems design. In A. diSessa, C. Hoyles, R. Noss, & L. D. Edwards (Eds.), *Computers and exploratory learning* (pp. 15–29). Berlin, Germany: Springer.
diSessa, A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and Instruction, 22*(3), 293–331.
English, L. D. (2010). Young children's early modelling with data. *Mathematics Education Research Journal, 22*(2), 24–47.
English, L. D. (2011). Data modelling in elementary and middle school classes: A shared experience. In T. De Silva Lamberg and L.R. Weist (Eds.), *Transformative mathematics: Teaching and learning.* Proceedings of the 33rd annual conference of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 250–258). Reno, Nevada, October 20–23.
English, L. D. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics Education, 81*(1), 15–30.
English, L. D. (2013). Reconceptualising statistical learning in the early years. In L. D. English & J. Mulligan (Eds.), *Reconceptualising early mathematics learning* (pp. 67–82). Dordrecht: Springer.
English, L. D. (2014a). Statistics at play. *Teaching Children Mathematics, 21*(1), 36–45.
English, L. D. (2014b). Promoting statistical literacy through data modelling in the early school years. In E. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 441–458). Netherlands: Springer.
English, L. D., & Watson, J. M. (2015). Exploring variation in measurement as a foundation for statistical thinking in the elementary school. *Journal of STEM Education, 2*(1), 1–20.
English, L. D., & Watson, J. M. (2016). Development of probabilistic thinking in fourth grade. *Journal for Research in Mathematics Education, 47*(1), 28–62.
Fielding-Wells, J., & Makar, K. (2015). Inferring to a model: Using inquiry-based argumentation to challenge young children's expectations of equally likely outcomes. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 89–115). Minneapolis: Catalyst Press.
Gavin, J. (2002). How much traffic? Beep! Beep! Get that car off the numberline! In R. Lehrer & L. Schauble (eds.), Investigating real data in the classroom: Expanding children's understanding of mathe and science. (pp 39–47). New York: Teachers College Press.
Gould, R. (2004). Variability: One statistician's view. *Statistics Education Research Journal, 3*(2), 7–16.
Greeno, J., & Hall, R. (1997). Practicing representation: Learning with and about representational forms. *Phi Delta Kappan, 78*(5), 361–367.
Hall, R., Stevens, R., & Torralba, A. (2002). Disrupting representational infrastructure in conversations across disciplines. *Mind, Culture, and Activity, 9*(3), 179–210.
Hall, R., Wright, K., & Wieckert, K. (2007). Interactive and historical processes of distributing statistical concepts through work organization. *Mind, Culture, and Activity, 14*(1–2), 103–127.
Hammerman, J. K., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17–41.
Hancock, C., Kaput, J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*(3), 337–364.
Hanner, S., James, E., & Rohlfing, M. (2002). Classification models across grades. In R. Lehrer & L. Schauble (Eds.), *Investigating real data in the classroom* (pp. 99–117). New York: Teachers College Press.
Hawkins, A. S., & Kapadia, R. (1984). Children's conceptions of probability: A psychological and pedagogical review. *Educational Studies in Mathematics, 15*(4), 349–347.
Horvath, J. K., & Lehrer, R. (1998). A model-based perspective on the development of children's understanding of chance and uncertainty. In S. P. LaJoie (Ed.), *Reflections on statistics: Agendas for learning, teaching and assessment in K-12* (pp. 121–148). Mahwah, NJ: Lawrence Erlbaum Associates.
Hutchins, E. (2012). Concepts in practice as sources of order. *Mind, Culture, and Activity, 19*(3), 314–323.

Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics Teaching in the Middle School, 5*(4), 240–246.

Kahn, J., & Hall, R. (2016, April). Getting personal with big data: Stories with multivariable models about global health and wealth. Paper presented at the American Education Research Association Annual Meeting, Washington D.C.

Knorr Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6,* 59–98.

Konold, C. (2007). Designing a data analysis tool for learners. In M. C. Lovett & P. Shah (Eds.), Thinking with data. (pp. 267–291). New York: Taylor and Francis.

Konold, C., Finzer, W., & Kreetong, K. (2015, July). *Modeling as a core component of structuring data*. Paper presented at the Ninth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL9). Paderborn, Germany.

Konold, C., & Harradine, A. (2014). Contexts for highlighting signal and noise. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen: Using tools for learning mathematics and statistics* (pp. 237–250). Wiesbaden, Germany: Springer.

Konold, C., Higgins, T., Russell, S.-J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education,* 2(1). Retrieved from http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art1/.

Konold, C., & Lehrer, R. (2008). Technology and mathematics education: An essay in honor of Jim Kaput. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed., pp. 49–72). Philadelphia: Taylor & Francis.

Konold, C., & Miller, C. D. (2011). *TinkerPlots: Dynamic data exploration [computer software, Version 2.2]*. Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Kosara, R., & Mackinlay, J. (2013). Storytelling: The next step for visualization. *IEEE Computer, 46*(5), 44–50.

Latour, B. (1999). *Pandora's hope: Essays on the reality of science*. Cambridge, MA: Harvard University Press.

Lavigne, N. C., & Lajoie, S. P. (2007). Statistical reasoning of middle school children engaged in survey inquiry. *Contemporary Educational Psychology, 32*(4), 630–666.

Lehrer, R. (2011, April). *Learning to reason about variability and chance by inventing measures and models*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Orlando, FL.

Lehrer, R. (2015, July). *Developing practices of model-based informal inference*. Paper presented at the Ninth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL9). Paderborn, Germany: University of Paderborn.

Lehrer, R., & Kim, M. J. (2009). Structuring variability by negotiating its measure. *Mathematics Education Research Journal, 21*(2), 116–133.

Lehrer, R., Kim, M.-J., Ayers, E., & Wilson, M. (2014). Toward establishing a learning progression to support the development of statistical reasoning. In J. Confrey & A. Maloney (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 31–59). Charlotte, NC: Information Age Publishers.

Lehrer, R., Kim, M. J., & Jones, S. (2011). Developing conceptions of statistics by designing measures of distribution. *International Journal on Mathematics Education (ZDM), 43*(5), 723–736.

Lehrer, R., Kim, M., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in modeling and measuring variability. *International Journal of Computers for Mathematics Learning, 12*(3), 195–216.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14*(1), 69–108.
Lehrer, R., & Schauble, L. (2000). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning, 2*(1–2), 49–72.
Lehrer, R., & Schauble, L. (2002a). Symbolic communication in mathematics and science: Co-constituting inscription and thought. In E. D. Amsel & J. Byrnes (Eds.), *Language, literacy, and cognitive development: The development and consequences of symbolic communication* (pp. 167–192). Mahwah, NJ: Lawrence Erlbaum Associates.
Lehrer, R., & Schauble, L. (Eds.). (2002b). *Investigating real data in the classroom: Expanding children's understanding of math and science*. New York: Teachers College Press. (Spanish Translation, Publicaciones M.C.E.P., Sevilla, Spain).
Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 635–679.
Lehrer, R., & Schauble, L. (2005). Developing modelling and argument in the elementary grades. In T. Romberg, T. Carpenter, & F. Dremock (Eds.), *Understanding mathematics and science matters* (pp. 29–53). Mahwah, NJ: Lawrence Erlbaum.
Lehrer, R., & Schauble, L. (2007). Contrasting emerging conceptions of distribution in contexts of error and natural variation. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 149–176). New York: Taylor & Francis.
Lehrer, R., & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education, 96*(4), 701–724.
Lehrer, R., Schauble, L., Carpenter, S., & Penner, D. E. (2000). The inter-related development of inscriptions and conceptual understanding. In P. Cobb, E. Yackel, & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design* (pp. 325–360). Mahwah, NJ: Lawrence Erlbaum.
Lehrer, R., Schauble, L., & Lucas, D. (2008). Supporting development of the epistemology of inquiry. *Cognitive Development, 23*(4), 512–529.
Lucas, D., Broderick, N., Lehrer, R., & Bohanan, R. (2005). Making the grounds of scientific inquiry visible in the classroom. *Science Scope, 29*(3), 39–42.
Makar, K. (2014). Young children's explorations of average through informal inferential reasoning. *Educational Studies in Mathematics, 86*(1), 61–78.
Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning, 16*(1), 1–24.
Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 152–173.
Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.
Manor Braham, H., & Ben-Zvi, D. (2015). Students' articulations of uncertainty in informally exploring sampling distributions. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 57–94). Minneapolis: Catalyst Press.
Manz, E. (2012). Understanding the co-development of modeling practice and ecological knowledge. *Science Education, 96*(6), 1071–1105.
Manz, E. (2015). Examining evidence construction as the transformation of the material world into community knowledge. *Journal of Research in Science Teaching, 53*(7), 1113–1140.
McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics, 45*(1–3), 103–129.
Metz, K. E. (1998). Emergent ideas of chance and probability in primary-grade children. In S. P. Lajoie (Ed.), Reflections on statistics. Learning, teaching and assessment in grades K-12. Mahwah, NJ: Lawrence Erlbaum Associates.
Moore, D. S. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.
Pellegrino, J. W., Wilson, M. R., Koenig, J. A., & Beatty, A. S. (Eds.). (2014). *Developing assessments for the next generation science standards*. Washington, DC: National Academy of Sciences.

Petrosino, A., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*(2–3), 131–156.
Pfannkuch, M., Budgett, S., Fewster, R., Fitch, M., Pattenwise, S., Wild, C., et al. (2016). Probability modeling and thinking: What can we learn from practice? *Statistics Education Research Journal, 15*(2), 11–37.
Porter, T. M. (1986). *The rise of statistical thinking 1820–1900*. Princeton, NJ: Princeton University Press.
Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education, 31*(5), 602–625.
Prodromou, T., & Pratt, D. (2006). The role of causality in the co-ordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal, 5*(2), 69–88.
Putz, A. (2002). How children organize and understand data. In R. Lehrer & L. Schauble (Eds.), *Investigating real data in the classroom* (pp. 27–38). New York: Teachers College Press.
Reading, C., & Reid, J. (2010). *Reasoning about variation: Rethinking theoretical frameworks to inform practice*. Paper presented at the Eighth International Conference on Teaching Statistics, Ljubljana, Slovenia.
Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201–226). Dordrecht, The Netherlands: Kluwer Academic Publishers.
Ridgway, J. (2015). Implications of the data revolution for statistics education. *International Statistical Review, 84*(3), 528–549.
Rodgers, J. L., & Beasley, W. H. (2014, October). *Using graphs to tell research stories, from many different perspectives*. Paper presented at the annual meeting of the Society of Multivariate Experimental Psychology, Nashville, TN.
Rosling, H. (2010, July). *What showbiz has to do with it*. Plenary paper presented at the International Conference on the Teaching of Statistics (ICOTS8), Ljubljana, Slovenia.
Rosling, H., Ronnlund, A. R., & Rosling, O. (2005). New software brings statistics beyond the eye. In E. Giovannini (Ed.), *Statistics, knowledge and policy: Key indicators to inform decision making* (pp. 522–530). Organization for Economic Co-Operation and Development.
Rubel, L. H., Lim, V. Y., Hall-Wieckert, M., & Sullivan, M. (2016). Teaching mathematics for spatial justice: An investigation of the lottery. *Cognition and Instruction, 34*(1), 1–26.
Rubin, A., Bruce, B., & Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 314–319). Voorburg, The Netherlands: International Statistical Institute.
Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257–270.
Schwartz, D. L., Goldman, S. R., Vye, N. J., & Barron, B. J. (1998). Aligning everyday and mathematical reasoning: The case of sampling assumptions. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 233–273). Mahwah, NJ: Lawrence Erlbaum.
Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics, 16*(6), 1139–1148.
Thompson, P. W., Liu, Y., & Saldanha, L. A. (2007). Intricacies of statistical inference and teachers' understandings of them. In M. Lovett & P. Shaw (Eds.), *Thinking with data* (pp. 207–231). Mahwah, NJ: Erlbaum.
Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
Tufte, E. R. (1997). *Visual explanations*. Cheshire, CT: Graphics Press.
Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. New York: Routledge.
Watson, J. M., & English, L. D. (2015). Introducing the practice of statistics: Are we environmentally friendly? *Mathematics Education Research Journal, 27*(4), 585–613.

Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*(1), 44–70.
Whitin, D. J., & Whitin, P. E. (2011). *Learning to read numbers: Integrating critical literacy and critical numeracy in K-8 classrooms*. New York: Routledge.
Wild, C. (1994). Embracing the "wider view" of statistics. *The American Statistician, 48*(2), 163–171.
Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal, 5*(2), 10–26.
Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–248.

Chapter 8
Learning About Statistical Inference

Katie Makar and Andee Rubin

Abstract This chapter reviews research on the learning of statistical inference, focusing in particular on recent research on informal statistical inference. The chapter begins by arguing for the importance of broader access to the power of statistical inference—which, until recently, has been accessible only to those with extensive knowledge of mathematics—and then traces the philosophical roots of inference. We outline the challenges that students have encountered in learning statistical inference and strategies to facilitate its learning that have capitalized on technological advances. We describe the emergence of informal statistical inference and how researchers have framed the idea over the past decade. Rather than consider formal and informal statistical inference dichotomously, we highlight a number of dimensions along which approaches to statistical inference may differ, providing a richer perspective on how formal and informal statistical inference are related. Cases from classroom research aimed at primary, secondary, and tertiary levels are used to illustrate how informal statistical inference has shaped new ways to approach the teaching and learning of statistical inference. Finally, we outline gaps in research on statistical inference and present our speculations on its future in light of new research on statistical modeling and big data.

Keywords Statistical inference • Informal statistical inference • EDA • Randomization • Simulation

---

K. Makar (✉)
School of Education, The University of Queensland, Brisbane, QLD, Australia
e-mail: k.makar@uq.edu.au

A. Rubin
TERC, Cambridge, MA, USA
e-mail: andee_rubin@terc.edu

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_8

8.1 Introduction

The focus of this chapter is on the development of educational approaches to statistical inference. In particular, we discuss efforts to help students understand formal statistical inference, challenges to these approaches, and the development of informal statistical inference as one response to these challenges. In Sect. 8.2, we introduce statistical inference as an idea and discuss the importance of providing students with access to the power of statistical inference. We outline philosophical roots of inference and how teaching and learning of statistical inference can be informed by a philosophical perspective. In Sect. 8.3, we include a brief overview of research on the learning of formal statistical inference and then focus Sect. 8.4 on the emergence of key ideas of informal statistical inference that have come out of the literature in the past 10–20 years. We use case studies from the literature in Sect. 8.5 to highlight these issues and then conclude in Sect. 8.6 by discussing the direction of the field, emerging work, and recommendations for further research.

8.2 Statistical Inference: The Heart and Power of Statistics

Inference is at the heart of statistics, as it provides a means to make substantive evidence-based claims under uncertainty when only partial data are available. In this section, we begin with an overview of statistical inference and its role in the teaching and learning of statistics (for an introduction to statistical inference within the field of statistics, see Part I). Next, we argue for the necessity of broader access to the power of statistical inference than traditional approaches afford. This argument will set the context for later sections of the chapter where we outline the challenges and opportunities for statistical inference at both the university level and for younger students. This section closes with a discussion of the parallels between statistical inference and philosophical inference. These parallels have been used to launch new ways of thinking about statistical inference, including ways to provide learners with access to its power prior to or without more formal statistical methods.

8.2.1 What Is Statistical Inference?

Where descriptive statistics give us specific knowledge (e.g., the typical height of children in this class is 124 cm tall), inference gives us general knowledge (e.g., based on this sample of students, we estimate the average height of 7-year-old children in the UK to be around 124 cm). Harradine, Batanero, and Rossman (2011) describe statistical inference as “the process of assessing strength of evidence concerning whether or not a set of observations is consistent with a particular hypothesized mechanism that could have produced those observations” (p. 235).

More broadly, Cobb and Moore (1997) describe statistical inference as "methods for drawing conclusions from data about the population or process from which the data are drawn" (p. 813). Both of these characterizations recognize that statistical inference can be made both from a sample to a population and from a sample to a process (or mechanism) that produced the sample. In the latter case, the population may not actually exist at the time. For example, assessing a random sample of outputs from an assembly line assumes that the sample is representative of all outputs that have been and/or will be produced by the assembly line, within a given timeframe and assuming no malfunctions (see Frick, 1998, for a fuller discussion of inference to a mechanism). Researchers have sought to broaden meanings of statistical inference by examining the more colloquial meaning of the word "infer" and then to adapt these colloquial meanings to a statistical context. For example, Rossman (2008) noted that colloquial definitions of the word "infer" include not just the conclusion that was drawn but the evidence and reasoning on which the inference is made. He argued that "inference requires going beyond the data at hand, either generalizing the observed result to a larger group (i.e., population) or by drawing a more profound conclusion about the relationship between the variables" (p. 5). Rossman recognized the role of chance variability as "fundamental" (p. 6) to statistical inference. Makar and Rubin (2009) included these elements in their broad interpretation of statistical inference as a probabilistic (nondeterministic) generalization using data as evidence. Their interpretation articulated the uncertainty embedded in a statistical inference and recognized that the claim (generalization) being made goes beyond the data available and is explicit about the evidence (data) used to justify the inference. Researchers have also emphasized the reasoning that leads to a statistical inference (often called inferential reasoning) in their expanded interpretation of statistical inference (e.g., Garfield & Ben-Zvi, 2008; Zieffler, Garfield, delMas, & Reading, 2008). These ways of characterizing statistical inference provide for a range of approaches and perspectives, including both formal and informal statistical inference, a distinction we elaborate on in Sect. 8.4.

Bakker, Kent, Derry, Noss, and Hoyles (2008) contrasted traditional approaches to hypothesis testing with inferences that are made in the workplace, using techniques such as statistical process control. They claimed that the concept of statistical inference has meaning beyond a final statement about a population, broadly referring to inference:

> in its general sense of drawing conclusions, including the possibly tacit reasoning processes that precede and support the explicit inference from a premise to a conclusion, a prediction, or a conjecture. The term [inference] not only includes deduction and induction, but also abduction. Abduction is inference to an explanation, a method of reasoning in which a hypothesis is formed that may explain the data. (p. 132)

Bakker and his colleagues (2008) therefore characterized statistical inference as being embedded in reasoning and explanations about the context in which the inference was being applied. By turning the focus toward the *purpose* and *reasoning* in statistical inference, there are opportunities to reconnect to inference as a tool for understanding phenomena.

8.2.2 *The Importance of Access to the Power of Statistical Inference*

Statistical inference is where the power of statistics lies. Formal methods of statistical inference allow one to make a fairly precise estimate of the output of an entire orchard of apples based on a careful selection of just a few trees. Inference gives medical research a way to make decisions about which of two treatments may be more effective in a population by collecting data on a study of patients. Unfortunately, for those without a strong foundation in the mathematical aspects of statistics, formal methods of statistical inference can act as a gatekeeper for access to the power of statistics. Harradine et al. (2011) and others point to a multitude of concepts and experiences that are important for understanding statistical inference including using simulations to understand variability between samples, understanding the role of randomness in sampling, and facility with probability concepts and their links to sampling distributions. However, Bakker and his colleagues (2008) contend that statistical inferences need to be made, and *are* made, by those in the workplace who do not have a background in the foundational ideas put forth by Harradine and his colleagues. Therefore, access to inferential power must not only be available to those with a background in statistics but also extended, to some degree, to those who do not know—and do not plan to learn—the underlying concepts outlined by Harradine et al. (2011).

The importance of access to the power of statistical inference goes beyond the workplace. In everyday life, people make predictions about a population or process without having all of the data. They create estimates out of uncertain information and make decisions about the future based on what they know now. Recognizing the kinds of statistical inference that are needed in everyday contexts can help to develop better ways to acknowledge and improve access to reasoning from—and beyond—data.

Bakker and Derry (2011) described two ways in which school-level statistics education often fails to achieve its goals, by producing students with (1) *inert knowledge*, "knowledge that students have learned to reproduce but cannot use effectively" (p. 6), or (2) *atomistic knowledge*, separate, discrete bits of knowledge about individual statistical tools that students cannot relate to one another. The challenge, Bakker and Derry proposed, is how to *sequence and build concepts to improve coherence* and provide access to the power of statistics earlier. Bakker and Derry see a focus on statistical inference as one possible solution to these problems, providing a holistic approach to learning statistics that begins to address these challenges. Many countries are concerned about the decline in those choosing to study mathematics and statistics, an issue that has serious implications in a knowledge economy (e.g., Australian Academy of Science, 2006; (UK) Council for the Mathematical Sciences, 2004). Therefore, access to the power of statistics through inference could improve students' valuing of statistics, appreciation of its relevance, and motivation to engage in further study. How can we make the power of statistical inference more accessible? We seek some inspiration from philosophical perspectives on inference.

8.2.3 Philosophical Roots of Inference

Moore (1990) is often cited as distinguishing statistics from mathematics by saying that "data are not merely numbers, but *numbers with a context*" (p. 96, italics in original). Indeed, the purpose of statistics, and particularly statistical inference, is to learn something new about a context based on available (but limited) evidence. Philosophy has a similar aim in using inference generally to make claims based on limited knowledge. Philosophical inference is useful to study because it can highlight habits of mind and foundational ideas underpinning inferential thought, which may provide fresh insights into the nature of statistical inference (Hacking, 2006; see also Nilsson et al., Chap. 11). We focus on philosophical inference in the context of uncertainty (as opposed to the deterministic inference used in mathematical logic), using the work of John Dewey (1910), who wrote extensively on inference in reflective thought and in inquiry. We then turn to research in statistics education that has drawn on philosophy to extend understandings of inference.

8.2.3.1 John Dewey's Inference

Dewey (1910) described inference as reflective thought that relies on evidence to reach a conclusion. He elaborated that "the data at hand cannot *supply* the solution; they can only *suggest* it" (p. 12, emphasis added), such as the way that one might infer that it has rained overnight by examining the grass. While inferences cannot be made with certainty, inference is not mere whim either. Dewey suggested several characteristics of inquiry that can help to replace impulsive or undeveloped speculations, including "habits of suspending judgment till inferences have been tested by the examination of evidence" (p. 66), telling "where and how to seek such evidence" (p. 67), and acknowledging that the quality of an inference depends on the sample and/or cases and how they are selected (pp. 88–91). In this way, Dewey argued that evidence becomes *part of* an inference. However, he also acknowledged that the source of an inference may begin less formally:

> Suggestion is the very heart of inference; it involves going from what is present to something absent. Hence, it is more or less speculative, adventurous. Since inference goes beyond what is actually present, it involves a leap, a jump, the propriety of which cannot be absolutely warranted in advance, no matter what precautions be taken. … The suggested conclusion so far as it is not accepted but only tentatively entertained constitutes an idea. Synonyms for this are supposition, conjecture, guess, hypothesis, and (in elaborate cases) theory. (p. 75)

Dewey argued that inference thus encompasses both the creative insight, which may be speculative, and methods to develop evidence relating to this speculation. Dewey's notion of inference involved a "fruitful interaction" (p. 80) that synthesized "the movement toward the suggestion or hypothesis and the movement back to facts" (p. 81). Therefore, Dewey argued that, "The aim of inference is to terminate itself in an adequate judgment of a situation, and the course of inference goes on through a series of partial and tentative judgments" (p. 101) in a process of inquiry.

8.2.3.2 Philosophy of Inference in Statistics Education

What lessons for statistical inference can we draw from Dewey's philosophical discussion of inference? Dewey's point above that inference has both a creative function and one that requires careful valuing of evidence finds a parallel in exploratory and confirmatory approaches to statistical inference. Embracing exploratory forms as an important part of inference enables us to generate new meanings. But we must also engage in a process of inquiry that suspends judgment while we seek supporting evidence and methods of "where and how." Wild and Pfannkuch (1999) describe the skepticism that statisticians must use during statistical inquiry as they seek insight and evidence. Within statistics education research, there is renewed interest in reconnecting with these exploratory approaches to statistical inference, to build meaning-making and recognize the need for inquiry to develop critical evidence (e.g., Makar, Bakker, & Ben-Zvi, 2011).

In exploring the role of context in statistical inference, Bakker and his colleagues used philosophical writings to broaden tacit assumptions about inference. Their work drew on the philosophies of Brandom, Dewey, Pierce, and Polanyi to argue that one cannot consider an inference without recognizing the reasoning, explanation, and personal knowledge within the context of the inference (Bakker et al., 2008; Bakker & Derry, 2011; Ben-Zvi, Aridor, Makar, & Bakker, 2012; Makar et al., 2011). Situating a discussion of statistical inference in a philosophical perspective reminds us to avoid an artificial separation between statistical knowledge and the rich contextual knowledge needed to apply inference meaningfully (Bakker & Derry, 2011).

8.3 Key Ideas and Efforts to Teach Students Formal Statistical Inference

This section describes research on students' difficulties grasping the ideas and techniques of formal statistical inference. It highlights ongoing efforts to make the topic more accessible, especially approaches that take advantage of the significant advances in computation that have accrued over the past 20 years.

8.3.1 Types of Formal Statistical Inference

While all statistical inference is an attempt to draw a conclusion about events, quantities, or situations beyond the data at hand, statisticians generally distinguish between two different kinds of inference: sample to population and experiment to causation (Cobb, 2007). The diagram in Figure 8.1 illustrates the relationship

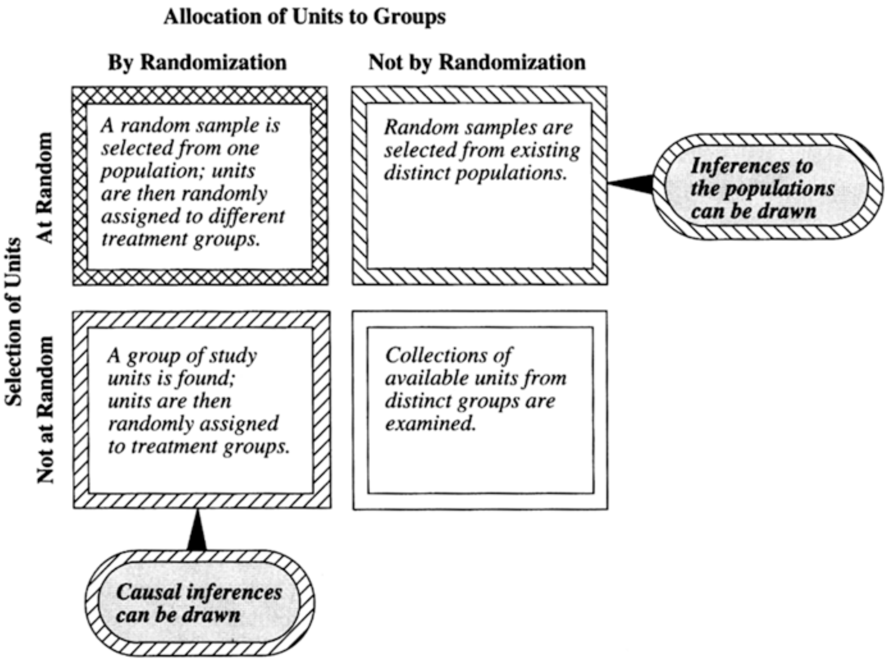


Fig. 8.1 Statistical inferences permitted by study designs (from Ramsey & Schafer, 2013, as reproduced in Cobb, 2007, p. 3)

between these two kinds of inference, based on the distinction between random selection of units and random allocation of units to groups.

Most of what students encounter are examples of the first type, in which a sample is drawn at random from a population and statistical techniques are used to figure out what can be said about the population. Inferences about differences between groups are possible in such designs, as long as all samples have been drawn randomly, but not about what might have caused such differences. The second type of inference is most common in medical and scientific settings, in which the effect of a particular treatment or intervention is being assessed. In these contexts, study units are not drawn at random, as they are often volunteers, but are assigned to treatments at random. This kind of design makes inferences about causes possible, but one must be cautious about generalizing beyond the sample. Both kinds of inference pose conceptual problems for students, although much more research has been carried out on difficulties students have understanding sample-to-population inference than experiment-to-causation inference (Pfannkuch, Budgett, & Arnold, 2015).

8.3.2 *Conceptual Difficulties in Statistical Inference*

Formal statistical inference is notoriously difficult for students to master, and many statistics education researchers have attempted to understand the roots of this struggle. Generally, researchers such as Castro-Sotos, Vanhoof, Van den Noortgate, and Onghena (2007, p. 99) have cited the need "for students to understand and connect many abstract concepts such as sampling distribution and significance level" as a reason for the widespread lack of understanding of statistical inference. Other researchers have identified more specific causes that suggest potential alternate approaches. Many authors pointed out the general difficulty people have with probabilistic reasoning (e.g., Kahneman, Slovic, & Tversky, 1982; Nickerson, 2004); Rossman (2008) quoted Stanovich (2007) as naming probabilistic reasoning as "the Achilles heel of human cognition" (p. 12). Rossman offered as an example the difficulty people have understanding statistical tendency, in contrast to the relative ease they have grasping a deterministic relationship. As Rossman put it, people "tend to ascribe deterministic explanations to chance phenomena and tend not to consider variability in general, and chance variation in particular" (p. 12).

8.3.2.1 Hypothesis Testing

Rossman also noted that the logical structure of Fisherian inductive inference is related to a particularly tricky kind of argument: modus tollens, the method of denying or law of the contrapositive. Modus tollens starts with a conditional proposition of the form "if P is true, then Q is true." This form of argument continues with the observation that the data indicate that Q is not true. The final step is to conclude from these two premises that, therefore, P is not true. A simple example of modus tollens: If a dog is a Dalmatian, it has spots. This dog does not have spots. Therefore, this dog is not a Dalmatian.

While the Dalmatian example seems simple, Rossman argued that less familiar examples of modus tollens reasoning can be quite challenging. Classical statistical inference has this form. We create a null hypothesis (P) and then claim that if it is true, some other proposition (Q) will also be true. When we have evidence that Q is not true, that allows us to also claim that the null hypothesis (P) is not true. However, statistical inference further complicates this kind of reasoning by throwing in "a probabilistic aspect … for good measure" (p. 13). It's not surprising that this form of statistical inference has proven to be a stumbling block for many students of statistics. There is a considerable body of research documenting students' difficulties understanding the structure of modus tollens and, consequently, interpreting *p*-values (e.g., delMas, 2004; Falk & Greenbaum, 1995).

Research has identified several additional aspects of the hypothesis-testing process that cause trouble for students. Some students conflate the null and alternative hypothesis or have trouble constructing an appropriate null hypothesis (Castro-

Sotos et al., 2007). Difficulties in understanding the meaning of a *p*-value are legion, with one common misinterpretation being that the *p*-value is the probability of the null hypothesis being true (Reaburn, 2014a). Castro-Sotos et al. (2007) have an extensive categorization of many of these misconceptions, further illustrating the complexity of the logic and foundational concepts within hypothesis testing.

8.3.2.2 Confidence Intervals

Partly in reaction to the conceptual difficulty of hypothesis testing (but also because of criticism of the approach from some statisticians, e.g., Harlow, Mulaik, & Steiger, 1997), confidence intervals have risen in popularity in the past two decades (Cumming, 2012; Wagenmakers, 2007). A confidence interval provides a range of parameter values that correspond to “plausible” populations from which the random sample might have been drawn, given a probabilistic definition of “plausible” (Reaburn, 2014b). Proponents of the use of confidence intervals point out advantages of this way of indicating the results of statistical inference over hypothesis testing, including: confidence intervals are in the same units as the point estimate; the width of the confidence interval indicates the precision of the estimate; and confidence intervals avoid the problematic logic of hypothesis testing.

While they may have certain advantages over hypothesis testing, confidence intervals are not transparent, either, and research has pointed out that they, too, are commonly misinterpreted by both students and scientists (Belia, Fidler, Williams, & Cumming, 2005). The majority of university students studied by delMas, Garfield, Ooms, and Chance (2007) at the end of an introductory statistics course thought that the level of confidence (e.g., 95%) indicated the percentage of all sample means that would lie within the confidence interval, rather than the probability that the confidence interval included the true mean. Reaburn (2014b) replicated this finding and described other common misconceptions about the confidence interval among tertiary students.

8.3.2.3 Additional Barriers to Understanding Statistical Inference

Other researchers have identified even more basic statistical concepts that are poorly understood by students (e.g., see Biehler et al., Chap. 4)—and, thus, impede their understanding of statistical inference. These include the concepts of distribution (Bakker & Gravemeijer, 2004), variation and covariation (Cobb, McClain, & Gravemeijer, 2003), sampling distribution (Saldanha & Thompson, 2002), sampling variability (delMas, Garfield, & Chance, 1999), and the idea that an aggregate value such as the mean or median can be used to compare distributions (Konold, Higgins, Russell, & Khalil, 2015; Konold & Pollatsek, 2002).

8.3.3 Attempts to Make Inference More Accessible

Given the documented difficulty of formal approaches to statistical inference—and concurrent with the increased availability of computational resources—statistics educators have come up with alternate approaches to inference that rely on being able to repeat a randomization procedure many times and keep track of the outcomes to determine the likelihood of a particular result. This method was described concisely by Cobb (2007) as "the three R's: randomize, repeat, reject" (p. 12). In slightly less telegraphic terms, his method can be elaborated to (1) randomize data production, (2) repeat by simulation to see what's typical, and (3) reject any model that puts your data in the tail. Cobb's "randomize" recommendation can be carried out in two different ways. In one approach, called bootstrapping, data are generated by repeated random resampling with replacement from a sample. In the other, called the randomization or permutation test, data are randomly reassigned to groups so that a comparison between groups can be made. Cobb's paper details a long list of reasons (e.g., the simplicity of the model and its match to the production process) why this approach is preferable to the more traditional approaches, which, he claimed, were the only option available before computers "made it possible to solve the problem directly" (p. 12). The randomization approach works equally well for the two kinds of inference described at the beginning of this section—sample to population (via bootstrapping) and experiment to causation (via the permutation test)—and it is becoming widely used in statistical practice (Hesterberg, Moore, Monaghan, Clipson, & Epstein, 2009).

Teaching about inference using randomization is a relatively new approach, but it has been enthusiastically embraced by some statistics educators and several curricula embodying its principles have been developed (Garfield, delMas, & Zieffler, 2012; Lock, Lock, & Morgan, 2012; Tintle et al., 2014). One group has developed a set of computer tools, called Visual Inference Tools (VIT) that support students in using visual reasoning to construct an understanding of the process of re-randomization and the inferences one can derive from it (Budgett, Pfannkuch, Regan, & Wild, 2013; Budgett & Wild, 2014; Wild, Pfannkuch, Regan, & Parsonage, 2017). Encouraging results have begun to appear as more people adopt Cobb's (2007) recommendations to use randomization methods as the basis of teaching inference. Budgett and Wild (2014), for example, reported that a course incorporating the VIT re-randomization module helped both university students and workplace students develop a basic understanding of the three R's process relatively quickly. They conjectured that these visual tools are especially effective when preceded by hands-on activities that demonstrate what the computer automates (i.e., randomly assigning group labels to the data points). In a cautionary note, the authors pointed out that even students who were able to carry out the randomization test fluently had trouble interpreting the tail proportion. In a related study using VIT, Pfannkuch and Budgett (2014) analyzed two university students' use of the tools to carry out both bootstrap confidence interval construction and the randomization test after an introductory course that used these tools. They noted that, in the context of the course, the tools appeared to facilitate students' development of statistical inferential concepts.

Tintle and his colleagues (Tintle et al., 2014; Tintle, Topliff, VanderStoep, Holmes, & Swanson, 2012; Tintle, VanderStoep, Holmes, Quisenberry, & Swanson, 2011) provided some of the only quantitative assessments of a randomization curriculum; their data showed that university students who were taught using randomization methods showed better understanding of basic statistical concepts than students taught with traditional methods. This improvement continued through a second test 4 months later, providing evidence of long-term understanding. Similar positive results were also found at different tertiary institutions with different faculty teaching the course. So, while research is still in the early stages, there is reason to be optimistic about the increased accessibility of ideas about statistical inference using a randomization approach.

8.4 Informal Statistical Inference as an Alternate Approach to Inference

Another approach to making statistical inference more accessible has drawn on informal approaches. As informal statistical inference has gained acceptance, its role has broadened from being primarily a path to formal statistical inference to include exploratory methods of analysis. In this section, we discuss the roots of informal statistical inference within exploratory data analysis, its emergence as a focus of research a decade ago, and shifts that have occurred within this research. We close this section with a discussion of several dimensions along which informal and formal statistical inference may differ.

8.4.1 Historical Background: Emergence from EDA

The focus on informal statistical inference was motivated by an impulse similar to the ideas that led to the creation of exploratory data analysis. Exploratory data analysis (EDA) was developed by John Tukey (1977) as a contrast to the more procedural confirmatory techniques that had dominated the field of statistical analysis until then. As Ainley and Pratt (2001) describe it, "EDA is concerned with organizing, describing, representing, and analyzing data, and makes extensive use of visual displays" (p. 5). Shaughnessy, Garfield, and Greer (1996) outlined the history of data handling, placing EDA at the center of a change in focus from statistical formula toward greater visualization, multiple forms of representation, and investigation of data. EDA provided approaches to exploring data without a need for sophisticated theoretical principles that relied on probability theory (Ben-Zvi & Arcavi, 2001; Prodromou & Pratt, 2006). It provided practical empirical tools that allowed a data investigator to visualize and explore data without assigning probabilities to findings. Part of Tukey's goal in developing EDA was to challenge the myth of simple linearity of the dominant confirmatory paradigm, where a question led seamlessly to a design, data collection, data analysis, and a clear answer (Tukey, 1980, p. 23). He claimed that this linear formulation of statistical inquiry hid the fact that neither the questions that launched the analysis nor the answer that finished it could be plucked from a context as a "tidy" (p. 24) bundle. Before a hypothesis could be written and confirmed, Tukey pointed out, a lot of digging was needed to find the insights in the data that were worth following up. He bemoaned the subsequent reduction of EDA to a few new data visualization tools: "Exploratory data analysis is an attitude, a flexibility, and a reliance on display, NOT a bundle of techniques, and should be so taught" (p. 23).

Although EDA has been available and widely accepted since the 1970s, technological advances in the past decade have resulted in more innovative tools for analysis based on visualization and simulation (Biehler, Ben-Zvi, Bakker, & Makar, 2013). These tools provide ways to manage more complex data sets, allowing students' statistical investigations to take on practices that more closely mirror the work of practicing statisticians (Wild & Pfannkuch, 1999). However, even as EDA was becoming more popular and easier to carry out, some statisticians and statistics educators expressed the concern that EDA loosened the connections between chance and data (Biehler et al., 2013; Pratt, 2011). On the one hand, EDA freed statistical analysis from the mathematical tedium of probability, but at the same time, it disconnected data analysis from the foundational concept of uncertainty. EDA-based explorations of data were often, therefore, descriptive in nature—they told the rich stories of the data that were collected but lost some of the power of statistics to extend beyond the data to uncertain claims about the population or process from which they were created. Informal statistical inference emerged as a way to build on the spirit of EDA while reclaiming the links between data and chance.

8.4.2 Emergence of Research on Informal Statistical Inference

Until three decades ago, research on statistical inference was largely focused at the university level. At that time, research took the standard university statistics curriculum as a given and studied how students reasoned (usually how their reasoning fell short) in the context of that curriculum. Konold (2007) noted five significant changes in the focus of statistics education research in the past 30 years, due to:

1. New understandings about student learning and children's capabilities
2. Concerns about meeting the statistical literacy needs of citizens now and in the future
3. Changes in school curricula, incorporating more experiences with data analysis from a young age

4. Technological tools that moved from number crunching to data visualization
5. Questioning the "core" ideas in statistics and probability

Konold argued that this reframing of core ideas has led to a revolutionary change in instructional design in statistical reasoning from developing "top down" to building "bottom up":

> Bottom-up instructional design … takes into account not only where we want students to end up, but also where they are coming from. Earlier approaches, in contrast, emphasized a top-down approach in which the college-level course—taken as the ultimate goal—was progressively stripped down for lower grades. … The objectives and content at a particular level are thus whatever was left over after subjecting the college course to this subtractive process. So grades 3–5 get line graphs and medians, grades 6–8 get scatterplots and means, and grades 9–12 get regression lines and sampling distributions. (p. 270)

The proliferation of work on informal statistical inference could be considered one result of this change in perspective. Once seen primarily as a method to facilitate high school and college students' transition from descriptive statistics to formal statistical inference (e.g., Garfield, Le, Zieffler, & Ben-Zvi, 2015; Zieffler et al., 2008), informal statistical inference is now commonly introduced to children and nonspecialist adults as a legitimate topic of its own (e.g., Bakker et al., 2008; Ben-Zvi, 2006; Makar, 2014; Meletiou-Mavrotheris & Paparistodemou, 2015). With these less technically oriented groups, the pedagogical aims are to develop coherence in their statistical ideas, give them earlier access to the power of statistical inference, and make connections between statistics and familiar contexts (Makar, 2016).

In response to the research described above, as well as influences from mathematics education, statistics education research conferences between 2000 and 2008 (especially the International Collaboration for Research on Statistical Reasoning, Thinking, and Literacy (SRTL) workshop but also the International Conference on Teaching Statistics (ICOTS) and the US Conference on Teaching Statistics (USCOTS)) began to include discussions of possible alternative approaches to teaching statistical inference. A shared pedagogical goal of these efforts was to harness students' informal reasoning strengths in constructing their understanding of formal statistical inference. Zieffler et al. (2008) posed the important question: "How can students' informal knowledge best be utilized in formal instruction [about statistical inference]?" (p. 42). Several groups of researchers have attempted to answer this question by engaging students in activities that focus on some critical aspects of statistical inference but strip away others. Some of the early frameworks for informal statistical inference include Ben-Zvi (2006); Makar and Rubin (2009); Pfannkuch (2006); Rubin, Hammerman, and Konold (2006); Rossman (2008); and Zieffler et al. (2008). While there are differences among these frameworks, several key common themes have emerged in describing informal statistical inference—claim beyond the data, expressed with uncertainty, use of data as evidence, consideration of the aggregate, and integration of context knowledge. Each of these is explained below.

Claim beyond the data. There is universal agreement that a key aspect of informal statistical inference is a claim that "goes beyond the data at hand" (Rossman, 2008, p. 5). Some authors have focused specifically on making claims “about unknown populations based on observed samples” (Zieffler et al., 2008, p. 44), and others have written more generally about “drawing conclusions from data … by comparing and reasoning from distributions of data” (Pfannkuch, 2006, p. 1). Pratt, Johnson-Wilder, Ainley, and Mason (2008) noted that teachers and learners are often unclear when they are talking about the sample at hand (which they call “Game 1”) and when they are talking about a larger unknown population (which they call “Game 2”). Many curriculum situations are actually examples of Game 1, as they ask students to reason only about data they have available (e.g., find the average height of students in our classroom). These activities could easily be extended to require students to consider what existing data mean about other as-yet-uncollected data, but few curricula take this step to Game 2.

Expressed with uncertainty. A second common theme in most descriptions of informal statistical inference is the presence of uncertainty. Statistical inference always involves some kind of probabilistic reasoning, since it comprises a claim about an unknown quantity, based on a known sample. In informal statistical inference, these expressions of uncertainty are not necessarily formal probabilistic statements but may be less precise or even qualitative. For younger students especially, stating that a sample is more likely to come from one population than another without quantifying the probabilities may be sufficient. Young students may only be able to distinguish a few levels of uncertainty and have trouble assigning any numerical values to them, but as they gain experience, their ability to recognize more subtle distinctions in uncertainty increase (Ben-Zvi et al., 2012).

Use of data as evidence. Makar and Rubin (2009) identified as a third component of informal statistical inference the key role of data as evidence for a claim. Being able to use data as evidence is a skill that develops over time, and young learners may have difficulty figuring out what the data they have collected mean about the question under investigation. Wild and Pfannkuch (1999) included recognition of the need for data as one of their central elements of statistical thinking, yet the connection between data and claims is often not made explicit for students (Fielding-Wells, 2010; Hancock, Kaput, & Goldsmith, 1992).

Consideration of the aggregate. Other authors have highlighted the critical role that aggregate measures such as the mean and median play in inference (Rubin et al., 2006). Inferences are made using aggregate measures of center, variability, shape, or covariation (Aridor & Ben-Zvi, in press), not individual data points. Reasoning about aggregates leads to consideration of signal (constant causes that are reflected in aggregate quantities) and noise (variable causes that introduce variability around a signal). Recent research (e.g., Konold et al., 2015) has reported that students often have trouble focusing on aggregate qualities of a distribution rather than on individual data points, so some efforts at increasing students’ understanding of inference have emphasized helping them to conceptualize aggregate measures.

Integration of contextual knowledge. Another key issue that has received attention in literature on informal inferential reasoning is the role of context. Langrall, Nisbet, Mooney, and Jansem (2011), for example, compared the level of statistical and inferential reasoning among students with different levels of contextual knowledge relevant to data they were analyzing. Their study showed evidence that students with knowledge about a context tended to provide greater depth of informal inferential reasoning than those without knowledge of the context. The importance of context knowledge has been highlighted by several studies of students' informal inferential reasoning (e.g., Dierdorp, Bakker, Eijkelhof, & van Maanen, 2011; Gil & Ben-Zvi, 2011; Madden, 2011). Makar and her colleagues (2011) explained that informal inferential reasoning is facilitated by an understanding of the problem context, access to statistical tools and concepts, and an inquiry-based environment.

8.4.3 Is It Formal or Informal Inference?

Since the introduction of the term "informal statistical inference" in the early 2000s, there has been considerable discussion about what makes a particular type of statistical reasoning "informal" or "formal." There is little question about some kinds of procedures; for example, the use of formulas or computation to carry out a *t*-test and determine a *p*-value is clearly an example of formal statistical inference. Similarly, most people would agree that young children claiming, based on the colored counters they have drawn out of a can, that they are pretty sure there are more red ones than blue ones is an example of informal statistical inference. But there are many examples that are much less clear-cut. For example, are randomization tests formal or informal? The perspective we take in this chapter is that there is actually a continuum of approaches to inference ranging from clearly informal to clearly formal, with many gradations in between. In taking this position, we are also stating that debates about whether particular approaches are formal or informal are not useful; rather, we should be studying what range of approaches are most beneficial for helping students appreciate and master the power of statistical inference. Garfield and Ben-Zvi (2008) take a similar approach in suggesting a set of activities that range along the formal-informal continuum to build students' understanding of inference.

One attempt to get more specific about "informality" was made by Zieffler et al. (2008), who described informal knowledge in the context of informal statistical inference as a combination of two kinds of information: (1) knowledge gained outside of school from life experience and (2) "less formalized" versions of topics gained from prior instruction (p. 42). This conception is useful in pointing out that informal knowledge may come from either life or school—but it is also somewhat recursive, as it describes informal primarily in contrast to formal. It leaves open the question of determining when reasoning crosses the line from formal to informal.

Rather than trying to answer this question by defining a specific point at which formal become informal, we take the position that it may actually not be helpful to label approaches as formal or informal. Instead, we propose a number of dimensions along which the complexity of statistical inferential reasoning—and the ways in which students learn about it—tends to vary. While the five dimensions listed below are themselves described as dichotomous, we see them as continua with many points between the extremes articulated. In addition, any individual approach to teaching about statistical inference is actually a set of choices, each of which may lie at a different point on the continua. For example, one approach might favor quantitative over qualitative description while relying on visualization over formulas:

Quantitative/qualitative: Many statistical inference methods quantify the likelihood of an outcome, as they involve deriving an explicit probability that the observed result could have happened by chance. Other, more qualitative approaches may involve judgments that an observed result is "surprising" or "unlikely" or that one result is more or less likely than another, without the assignment of numerical values to these probabilities.

Closed-form formulas/simulation: Traditional statistical inference is based on closed-form formulas, made possible by simplifying assumptions about underlying theoretical distributions. Newer approaches take advantage of computational power to estimate the probability of a result through repeated simulations, an approach that does not require the same simplifying assumptions that traditional methods do. These computationally based methods are not derived using traditional algebraic manipulation, yet they are becoming more accepted as standard statistical practice.

Diversity of images of distributions: Many textbook explanations of traditional statistical inference rely on a single iconic image of a theoretical standard normal distribution, with mean and standard deviation indicated (0 and 1, respectively) and 0.05 and 0.01 tails shaded. Other approaches tend to have more pictures, illustrating empirical data distributions, often non-normal, with superimposed visualizations such as box plots to help highlight the shape, central tendency, and variability of these data.

Choices of measures of central tendency and variability: To take advantage of mathematical methods available for the normal distribution, traditional statistical inference used mean and standard deviation as measures of central tendency and variability, respectively. Approaches to statistical inference using measures such as median and IQR (which do not assume that a distribution is normal) are becoming more widely accepted.

Community acceptance: Traditional formal statistical inference is accepted as valid and official in the statistical and scientific communities, even though it is recognized that an idealized normal distribution model is often not a good match to empirical data. Informal approaches to statistical inference are sometimes more idiosyncratic and individual and may not be accepted in the wider community. Thus, some approaches that are currently viewed as informal because they are not part of the mainstream may become "more formal" as the statistical, scientific, and educational communities shift in their adoption of inferential techniques.

In the next section, we describe several examples of teaching and learning informal statistical inference, which themselves vary along several of these dimensions.

8.5 Illustrations of Informal Statistical Inference

In this section, we describe cases of teaching and learning to illustrate key findings in developing inferential thinking through informal statistical inference. Because students at different levels are expected to apply different kinds of reasoning, we include separate cases that reflect thinking at the primary school, secondary school, and tertiary levels. Our secondary school example highlights teachers' reasoning as an example of the kinds of insights that are accessible to students in middle and high school.

8.5.1 Primary School Example of Informal Statistical Inference: Experiencing Powerful Statistical Ideas

The aim in primary school is not to move children toward formal statistical inference, but rather to capitalize on their everyday experiences with predictions, provide access to powerful statistical ideas, and create coherent opportunities to develop statistical reasoning (Makar, 2016). Although they have not yet learned descriptive statistics, it is beneficial "to elaborate the conceptual struggle that needs to take place for young students to engage in inferential reasoning" (Pratt et al., 2008, p. 108). We illustrate this with a case study (Fielding-Wells & Makar, 2015) from a Year 3 (aged 7–8 years) primary mathematics class in Australia to highlight how key ideas from informal statistical inference appeared in students' reasoning—making a claim beyond the data, expressing the claim with uncertainty, using data as evidence for the claim, working with data as an aggregate (distribution), blending chance and data, and integrating statistical reasoning with contextual knowledge.

In this study, the children were responding to the question, "What is the best card for winning addition bingo?" Addition bingo is played by placing all combinations of the sum of two numbers (from 1 to 10) on individual slips in a box and drawing them out one by one. Children mark off each sum if it appears on their 5 × 5 bingo card. For example, if the teacher drew out "4 + 7," then children crossed off one (and only one) occurrence of 11 if it appeared on their card. To investigate their reasoning, children were asked to fill in numbers for the 25 spaces of a blank bingo card to try and maximize their chances of winning (crossing off all 25 numbers on their card).

As the children played the game, they kept track of the numbers that were pulled out of the box and used these data to investigate strategies for selecting numbers for their card that would increase their chances of winning the game. The game therefore allowed them to make informal statistical inferences (Makar & Rubin, 2009) within a practical context because (1) their predictions (in the form of an addition bingo card) went beyond the data generated in each game—that is, the prediction aimed to create a card that would apply to current and future plays of the game; (2) most predictions (beyond the first game) were based on their experience with data from the game as evidence; and (3) predictions were expressed with uncertainty as they could not be certain their card would win. Although students initially considered individual numbers when playing the game, they began to realize the utility of using a distribution (aggregate) as a tool for selecting the numbers to put on their bingo card.

Addition bingo is a different context to learn probability than working with coins, dice, and spinners because unlike these contexts, the underlying distribution is unknown to the children. That is, there was no way for them to "check" if their solution was correct; the quality of their response depended on their reasoning to justify their conclusions with evidence. Although the formal terminology was not explicitly introduced, the game created opportunities for students to build a sample space, create and compare theoretical and empirical distributions, calculate probabilities of outcomes, and articulate informal inferential reasoning.

In their first attempt at creating a winning addition bingo card, most children expected outcomes of the game to be equally likely. Although this was not the best model for the data, these expectations were based on informal inferential reasoning because they represented the students' predictions beyond their own cards to the process that generated the data for every game. Children listed the numbers 2–20 on their cards plus a few "lucky" numbers to fill in the extra spaces. This equiprobability bias was challenged when the game was played, as several outcomes had multiple frequencies (e.g., there are five possible slips with a sum of 6), while others never appeared at all. To investigate why numbers were appearing multiple times, the students employed a variety of strategies, including dumping the slips of outcomes from the box onto the floor and tallying them (Fig. 8.2, top left), writing the possible ways to generate each sum from 1 to 20 as a narrative (top right), filling in an addition chart (not shown) and then counting the frequency of each outcome, or organizing their findings on a number line (bottom). The teacher helped the class create an addition chart (forming a sample space) to discuss expected relative frequencies. This sample space graphed as a dot plot (which they called "Paul's mountain," Fig. 8.2, bottom) became a representation students adopted to describe the possible outcomes of the game.

Students often overestimated expected frequencies in subsequent iterations of play. In critiquing a peer's choice of putting four 16s on her card, one student, Jess, argued that 16 was unlikely to occur that often:

> Well, I'm not trying to be mean to Lorena or anything, but like 16 isn't a really popular number so it might not come out as much as four times. It could, but it sort of like, is only a possible chance of 16 coming up.

In Jess' articulation, she used informal inferential reasoning to claim that 16 was unlikely to be called out, emphasizing her claim with uncertainty: "it could, but it … is only a possible chance."

When the numbers pulled from the box in multiple games were combined on the board (Fig. 8.3), students noted how the data they collected differed from their expectation (a triangular shape like "Paul's mountain"). Although the children didn't understand the formal probability and statistics concepts behind the problem (e.g., sample space, expected values), they were using similar ideas at a level

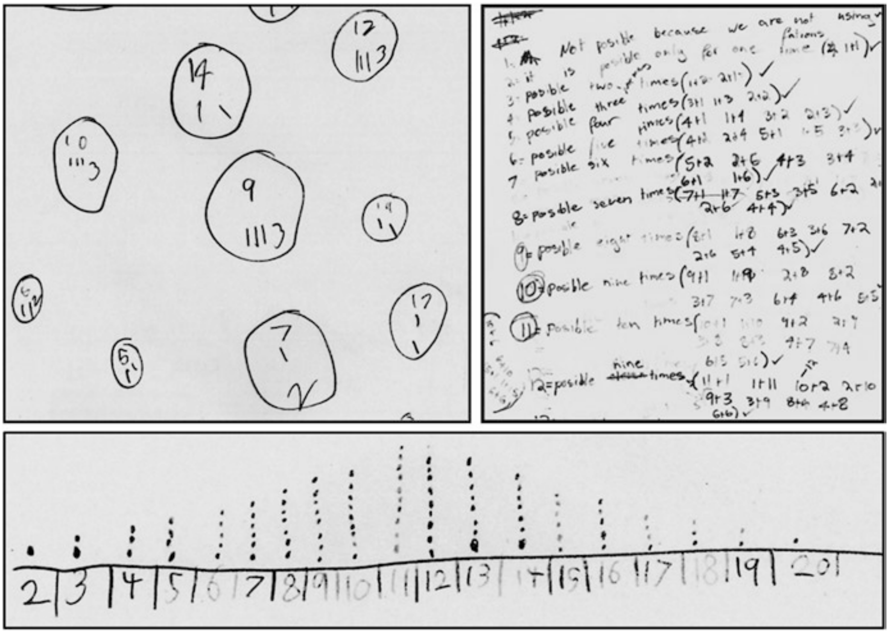


Fig. 8.2 Students' strategies to develop evidence for the frequencies of outcomes

appropriate to their age to articulate their informal inferential reasoning. Fielding-Wells and Makar (2015) cautioned that not all children reasoned from the distribution (aggregate) and some persisted with relying on "lucky" numbers in creating their addition bingo cards. These children were less successful in playing the game, and it was hoped that they would at some point in their schooling learn to trust the distribution to change their beliefs about luck and randomness.

Researchers outside of statistics have argued that young children should encounter powerful mathematical ideas, even if the children don't yet understand all of their details or implications of the ideas (e.g., Mulligan & Mitchelmore, 2013). The experiences in Fielding-Wells and Makar's (2015) study exposed children to problems that required informal statistical inference, providing them with opportunities to informally work with powerful statistical concepts and structures. The experience challenged the children's equiprobability bias (Lecoutre, Durand, & Cordier, 1990) and moved them toward meaningful, age appropriate exploration of distributions and aggregate reasoning. Furthermore, because the context was one that involved informal statistical inference, it integrated data and likelihood through informal versions of probability and statistical concepts—sample space, sampling variability, empirical and theoretical probability distributions, and calculations of probability—around a meaningful problem in coherent ways (Bakker & Derry, 2011).

8.5.2 *Middle/High School Example of Informal Statistical Inference: Using Aggregates to Support Inference*

The following example is from the Visualizing Statistical Relationships (ViSOR) research project, in which middle and high school teachers investigated how statistical visualization tools could support their own and their students' understanding of statistical reasoning (Rubin et al., 2006). It illustrates an emerging understanding of informal statistical inference that is accessible to middle and high school students using appropriate visualization tools, extending informal inferential reasoning beyond the primary grades example in Sect. 8.5.1 in several ways: in its explicit use of aggregate values, a recognition of the co-occurrence of signal and noise, and in relatively more sophisticated reasoning about variability. It also illustrates an example of inference from a sample to a process, rather than to a finite population.

A computer tool is central to this example: teachers used TinkerPlots (Konold & Miller, 2005) to explore data about variability in a process. The following story and accompanying data set were developed by Konold (2005):

- The Mus-Brush Company produces mushroom brushes, using a machine whose normal output is on average 215 brushes every 2 min. Output is recorded in terms of the unit "bptm," which stands for "brushes per two minutes."
- If the electricity to the machine is interrupted, even briefly, it will slow down, and the output of the machine will be 10% lower on average.
- The Mus-Brush Company was robbed last night; in forcing the door open, the thief disrupted the electricity, and the machine became less productive from that time on.
- There is a suspect who has an alibi between midnight and 3 AM, so the police want to know if the break-in occurred before midnight or after 3 AM, since the suspect has no alibi for those times.
- We have data on Mus-Brush production every 2 min from 8 PM until 6 AM. Our job is to decide whether there is enough evidence to argue that the break-in occurred between 12 and 3, thus freeing the suspect.

Figure 8.4 is a TinkerPlots graph that many teachers created early in their analysis by first plotting the case number, which corresponds to the order in which the measurements were collected, on the horizontal axis and the output (brushes per two minutes or bptm) on the vertical axis. They then divided the points into hour-long "bins," each of which contains all 30 data points for that hour; each bin is labeled by its hour, beginning with "eight" (PM) and continuing through "five" (AM). Within each bin, however, the points are no longer ordered by time.

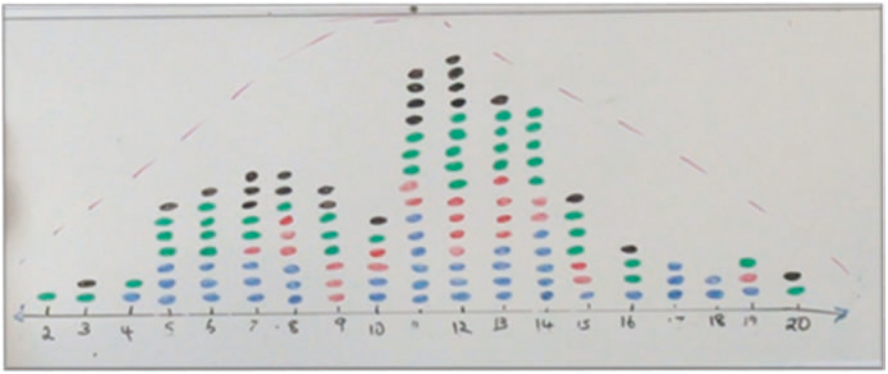


Fig. 8.3 Outcomes of multiple games combined on a number line (*dot plot*) with expectations of the shape of the data superimposed above it (*dotted line*)

To get a sense of trends in the data, many teachers colored the points on the graph according to their output value. Points that represented low output (less than 183 bptm) were colored blue, those with medium output were purple (from 183 to 229 bptm), and those with high output (greater than 229 bptm) were green. One pair of teachers noted, in examining this graph, that there were no high outputs in hours three (from 3:00 to 4:00 AM) and four (from 4:00 to 5:00 AM) and that those 2 hours also had the most low outputs. This led them to hypothesize that something happened between 2:00 and 3:00 in the morning, just before the hours in which they saw lower output.

These first two analysis moves are interesting in that teachers created representations that ignored some of the details of the data. Grouping the data into bins created ten distributions rather than a single time series, and coloring the data by high/medium/low created three categories rather than a range of individual values. A plausible interpretation is that the teachers visualized the data in chunks in recognition of the inherent variability of the process, hoping to see general trends over time, rather than changes from one data point to the next. These moves could indicate an appreciation for the importance of both signal and noise, as they present the data in a way that may make the signal more discernable.

To investigate their hypothesis that something happened during the hour from 2:00 to 3:00, these teachers added the means of each hour's production to the graph, looking for what they would consider a "significant" drop between one hour and the next. In Fig. 8.5, the mean output for each hour appears as a blue triangle to the left of the corresponding bin. For example, the mean production for the 8 o'clock hour is at the far left of the graph, around 220 bptm. On Fig. 8.5, teachers focused on the means of hours one and two and noted that there was a large drop between them. They reasoned that the machine was operating at full capacity during hour one but

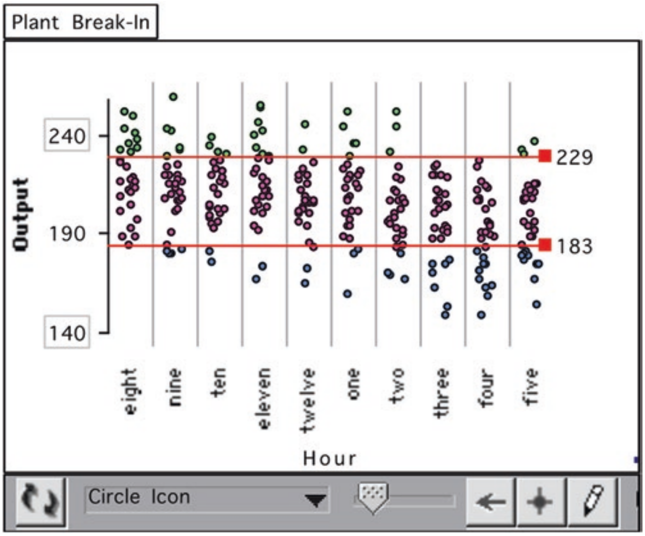


Fig. 8.4 Machine output separated into hour bins, colored by low, medium, and high output

began to operate at a reduced capacity during hour two and that, thus, the break-in occurred during hour two—i.e., before 3:00.

Other teachers used Fig. 8.5 to make a different argument. They noticed that the mean of hour two was 201 bptm, that of hour three was 196 bptm, and that of hour four was 191 bptm, a drop of 5 bptm each hour. This made it seem possible that the machine was operating normally until 3 AM and then began to operate at a reduced capacity during hour four. To decide whether it was more likely that the break-in occurred during hour two or during hour four, the teachers decided to look at the data in half-hour chunks, as in Fig. 8.6, to see if there was a more obvious change in mean production at this grain size. They argued that if the break-in occurred in the middle of some hour, the mean of the hour during which it happened would be an average of both normal operation and reduced operation and would thus not show the significant drop that would indicate a break-in.

In Fig. 8.6, one teacher found what she had hoped to see. The mean production between 2:00 and 2:30 was 207 bptm, reasonably close to the 215 "normal" value. But between 2:30 and 3:00, it dropped to 195 bptm. The arrows in Fig. 8.6 indicate these half-hour means. This change is larger than any other drop between two consecutive half hours, and the mean production continues to vary around 193 bptm after this time, about 10% less than normal production. The teacher therefore reasoned that the break-in most likely occurred around 2:30, at the boundary between the two half hours.

This series of graphs and the reasoning that accompanied them illustrate several important aspects of informal statistical inference. First, there is evidence of a deep

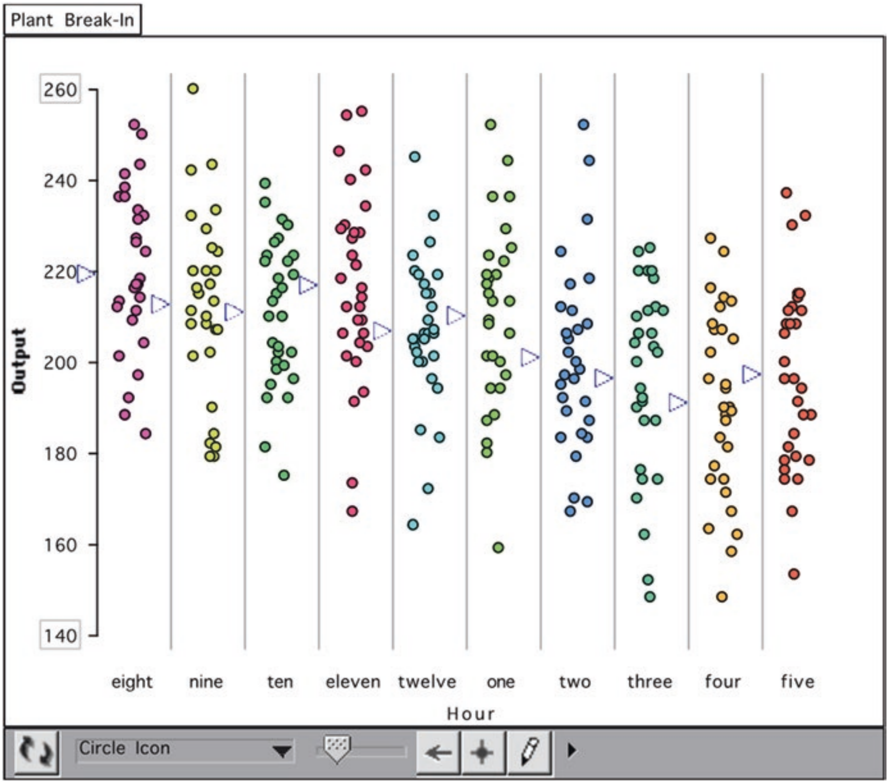


Fig. 8.5 Data (with means marked) of each hour's production

understanding of the ubiquity of variability. One teacher observed of Fig. 8.6, "The mean seems steady between 11:30 pm and 2 am," when, in fact, the mean varies between 206 and 217 during that time. We consider this plausible evidence of their comfort with variability, at least in the context of a process whose variation seems natural. Because the Mus-Brush context involves inference to a process rather than a population, it may be a context in which it is easier for students and teachers to see signal and noise as coexisting aspects of a phenomenon.

The teachers' reasoning also provides evidence that they recognized the utility of an aggregate measure of central tendency (in this case, the mean) as an indication of a signal in the presence of noise. TinkerPlots facilitated this use of the mean as an indication of signal in two ways. First, since the value of the mean was visually displayed on the graph, a pattern in means over time was relatively easy to notice. Second, it is easy in TinkerPlots to change the width of bins in a graph (as in going from Fig. 8.5 to 8.6), with the values of the bin means being updated automatically.

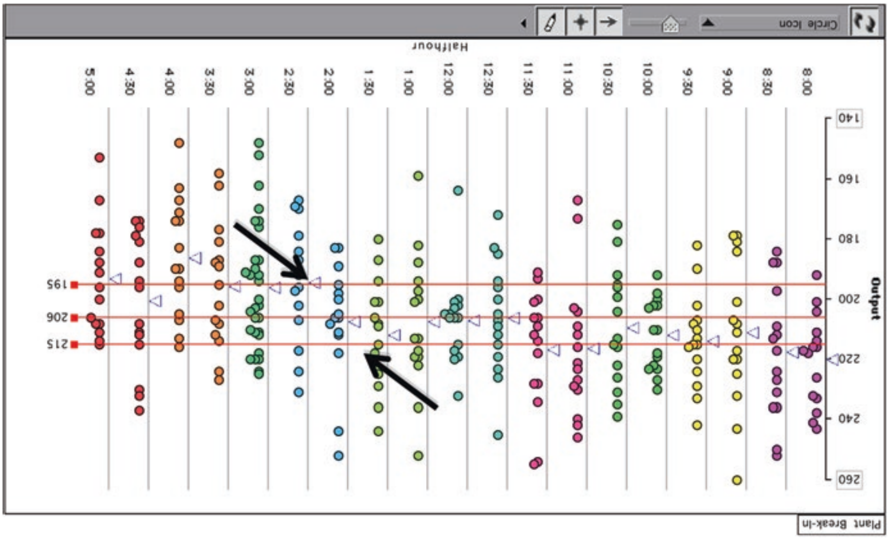


Fig. 8.6 Graph showing mean production in half-hour intervals

In sum, we note that, while these teachers did not measure variability in any quantitative way or carry out any significance tests, their understanding of the roles of variability, aggregate measures, and trends in statistical inference is statistically sophisticated and could provide a useful basis for more formal techniques.

8.5.3 Tertiary Examples of Informal Statistical Inference: Using Simulations and Randomization

Recently, informal approaches have been increasingly used at the tertiary level to improve students' grounding in statistical ideas prior to or in conjunction with more formal procedures (e.g., Garfield et al., 2015; Pfannkuch et al., 2015; Ramsey & Schafer, 2013). We therefore look at research on informal statistical inference in this section as coming full circle. That is, many of the studies currently undertaken on informal statistical inference at the university level do not begin with formal descriptive statistics, the central limit theorem and sampling distributions (top-down, Konold, 2007), but rather start with university students' conceptions of data and randomness (bottom-up) and develop their informal inferential reasoning from that basis. In this section, we briefly describe two cases that immerse students in informal aspects of statistical inference using simulations. The first is from the US CATALST group who use informal statistical inference and simulations to introduce hypothesis testing. The second is from a research group in New Zealand who use visualization tools to improve students' access to inference with randomization tests.

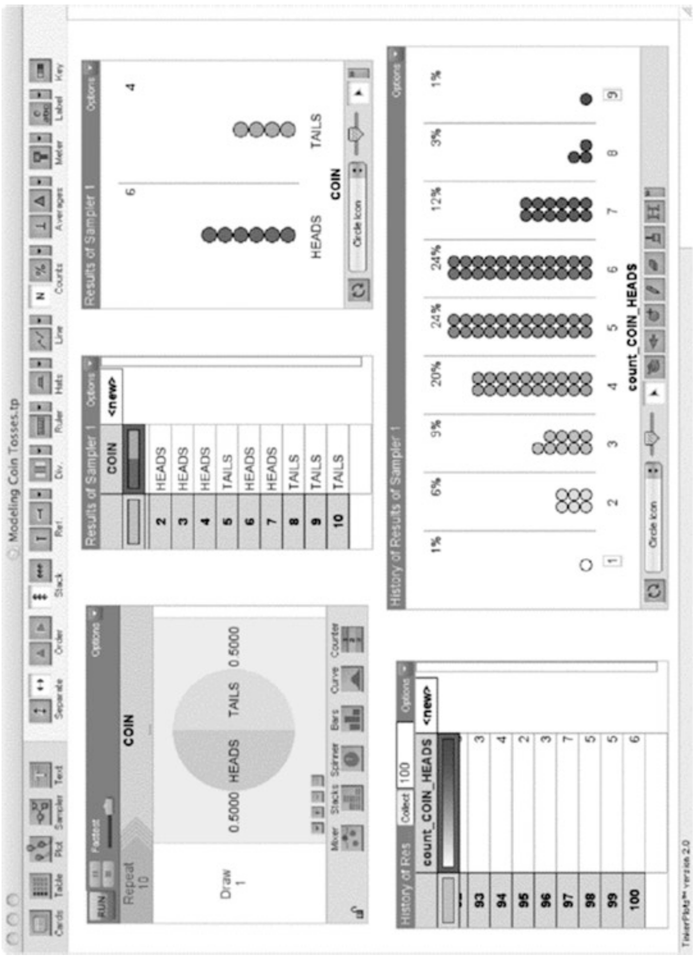


Fig. 8.7 Simulation in TinkerPlots of the number of heads obtained from tossing a coin 10 times, repeated 100 times (from delMas et al., 2014)

8.5.3.1 Tertiary Case 1: Using Simulations to Transition to Hypothesis Testing

Researchers from the CATALST group based their approach to developing statistical inference on a modeling perspective that incorporated simulation, resampling, and the core logic of inference (Cobb, 2007). Their aim was to develop students' initial understandings of the logic of inference by beginning with informal statistical inference. Prior to introducing formal methods of hypothesis testing, the introductory university course developed by CATALST engaged students in solving problems that required inferring from simple theoretical models and/or simulations. The course used activities that built on students' prior knowledge about sampling and sampling variability to generate potential models to solve and justify their solution to each problem (delMas, Garfield, & Zieffler, 2014; Garfield et al., 2012, 2015).

Throughout the course, students were asked to engage in the sequence—model, randomize and repeat, and evaluate—for problems of increasing complexity and formality. At each level, students used software to create a theoretical model (often after an initial experience with physical coins or dice), repeatedly collected random samples from the model, and then evaluated the "unusualness" of a particular outcome of interest (delMas et al., 2014). For example, students discussed whether it was reasonable to obtain ten heads from ten tosses of a fair coin. They tested their prediction by simulating 100 trials of the experiment and calculating the relative frequency of obtaining 10 out of 10 heads from the resulting empirical sampling distribution. Figure 8.7 shows an example of this simulation in TinkerPlots with the model of the coin (top left), table and graph of the outcomes of a single trial of 10 coins (center and top right), and, finally, the table and graph of the distribution of number of heads obtained from 100 trials of 10 coin tosses (bottom). In this simulation, there were no occurrences of 10 heads out of 10 coin tosses and just one occurrence of 9 heads in the 100 repetitions of 10 coin tosses (bottom right).

By assessing their speculation against the data distribution, students gained valuable insight into the utility of the model—and the data as evidence—for making claims beyond the data. In rerunning the simulation, they gained an informal sense of sampling variability, which allowed for important discussions about the difference between the population (theoretical model), a sample (10 coin throws), and a sampling distribution (100 sets of 10 throws). These initial informal experiences provided key contextual links to distinguish these concepts, ideas that research has documented as difficult (Castro-Sotos et al., 2007; Chance, delMas, & Garfield, 2004). The next unit of the CATALST course had students simulate the difference between treatment and control groups. The intent was to again draw on randomization to come to a conclusion and to introduce the concept of *p*-values—using informal approaches as an introduction to the formal procedures of significance testing. The final unit in the course supported students in hypothesizing a model from samples of data to emphasize the power of statistical inference to make claims about a population or process based on limited (sample) data generated from it.

8.5.3.2 Tertiary Case 2: Randomization and Bootstrapping Methods

University researchers in New Zealand have also adopted Cobb's (2007) recommendations, creating a large project to study how to introduce tertiary students to inferential ideas using bootstrapping and randomization methods (Budgett et al., 2013; Pfannkuch et al., 2015; Pfannkuch, Wild, & Regan, 2014; Wild et al., 2017). As part of this work, they developed the *Visual Inference Tools* software (VIT, www.stat.auckland.ac.nz/~wild/VIT), which provides dynamically linked graphs that track multiple runs of a simulation so that a user can see the distribution of results. The group considers their methods "partial informal inference," as they use a formal inferential method (the randomization test), which is part of the toolkit of professional statisticians, but do not introduce students to the formal ideas of null hypothesis, *p*-values, and significance.

Of particular interest to this research group is experiment-to-causation inference, since it is an often-overlooked aspect of statistical reasoning, but of the utmost importance in the study of statistics as they are actually used. The example illustrated below in Fig. 8.8 is the analysis of an experiment in which the efficacy of taking fish oil vs. regular oil on lowering blood pressure was tested. In the top graph of Fig. 8.8, the actual data distribution is shown, along with the difference of means of blood pressure in the two treatment groups (7.71 mmHg). The data were then randomly reassigned to the two groups, the difference in the resulting group means recorded (Fig. 8.8, middle graph), and the process repeated. The distribution resulting from 1000 repeated random reassignments is displayed in the bottom graph of Fig. 8.8, which allows students to judge whether the actual difference of 7.71 mmHg would be considered unlikely, i.e., how often a difference of 7.71 mmHg or greater would occur if the treatment had no effect. In this case, a difference that large in that direction would occur only 8 in 1000 times, so it is relatively unlikely.

Randomization methods such as these are highly visual and generally more accessible to students than conventional methods involving null hypotheses and *p*-values, therefore "increasing the accessibility of data exploration and inferential ideas to wider audiences" (Wild et al., 2017, p. 21). However, research with these tools has also demonstrated the complexity of the process of untangling the multiple aspects of uncertainty at the heart of making causality judgments in probabilistic settings. For example, in experiment-to-causation settings, students were more likely to invoke unrelated causal knowledge to explain differences than they were in sample-to-population inference. Students also had trouble figuring out how to apply ideas they had learned in the context of sample-to-population inference to experiment-to-causation inference, including notions of sample size, tail proportion, confounding variables, and generalization (Pfannkuch et al., 2015). The complexities of untangling these concepts suggest that "developing students' understanding of causality in a probabilistic setting will require multiple experiences over several years" (p. 21).

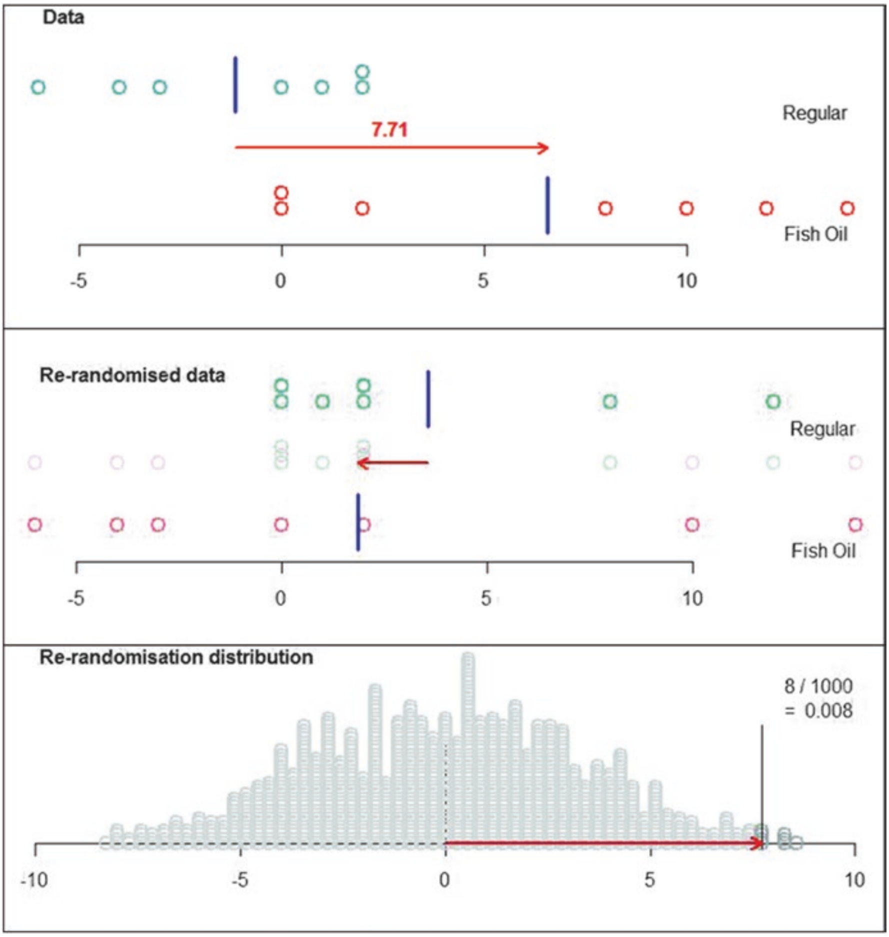


Fig. 8.8 Screenshot of dynamically linked plots in VIT randomization test (Pfannkuch et al., 2015)

8.5.4 Insights into Informal Statistical Inference from the Cases

The cases above highlight two key issues in engaging students in informal statistical inference. First, they emphasize the potential for informal statistical inference to help students across a number of levels (young children through university) to work productively with powerful statistical ideas before or instead of studying formal hypothesis testing. Second, the cases make it clear that separating informal and formal statistical inference dichotomously does not capture the utility of putting informal inferential reasoning into practice. While we consider them all examples of "informal" inference, there are aspects of formalism in them, especially as the students involved get older. In this sense, these cases blur the boundaries between formal and informal statistical inference and point instead to the value of focusing on students' inferential reasoning developed through visualization, simulation, and powerful problem contexts.

8.6 Summary and Future Directions

In sum, we have argued that statistical inference—formal, informal, or in-between—is where much of the power of statistics lies. Statistical inference, similar to Dewey's notions of inference, relies on evidence but goes beyond it from "what is present to something absent" (Dewey, 1910, p. 75). The ability to make statements or predictions beyond the data at hand, with the inclusion of uncertainty, provides students with the potential for insights they can use to win games, catch burglars, and test drugs. We have seen how the structure of traditional approaches to formal statistical inference can frustrate novices, and how new approaches to inference, collectively called "informal statistical inference," can provide students of all ages with more powerful statistical experiences. The four case studies included in this chapter illustrate what informal statistical inference might look like as students as well as teachers become more sophisticated in their statistical reasoning. They share a reliance on visual representations, use of technology specifically designed to illuminate statistical concepts (for middle school and beyond), and engaging research-based curriculum tasks.

In many ways, the study of informal statistical inference is quite young, and it is hard to predict where research will head next. From our perspective, however, there are a number of directions that we see as particularly likely to assume importance in the near future, either because they represent major changes in the statistical milieu or because they are natural outgrowths of current work. Here we briefly discuss five such areas of inquiry: (1) the relationship between statistical modeling and statistical inference, (2) data science and big data, (3) reconsiderations of probabilistic reasoning, (4) inference using time series and correlations, and (5) the potential of developing a statistics curriculum that develops students' informal inferential reasoning from kindergarten through tertiary levels. In each of these, we propose ideas for research that would extend our understanding of the development of inferential reasoning.

Statistical modeling has been studied for many years, including by those working with children (e.g., English, 2012; Konold, 1994; Konold, Harradine, & Kazak, 2007; Lehrer & Schauble, 2000, 2004). In most of these cases, statistical models are a means of making inferences about the contexts they are representing. There has been a strong resurgence in statistical modeling in the past few years, possibly as a result of new research in informal statistical inference and its shared goal of enhancing access to the power of statistical inference (see Lehrer & English, Chap. 7). As one indication of this connection, a special collection of papers on reasoning about models and modeling in the context of informal statistical inference was published in 2017 in the *Statistics Education Research Journal*.

A focus on modeling also provides an opportunity for collaboration between the math education and statistics education communities, as mathematical models may act as "boundary objects" that support conversation between the communities without requiring a single definition of "model" (Groth, 2015). Since, in either mathematics or statistics, modeling is a tool that allows us to understand empirical situations better, there are obvious overlaps between the research concerns of mathematics and statistics educators with respect to modeling.

The advent of "big data" has brought statistical reasoning into the spotlight, as people in a wide variety of fields seek insight from the flood of data newly available. Data science is touted as a driver of the economic future, and data skills have joined the ranks of abilities necessary for successful employment. However, much of the scholarship in this area is being done by computer scientists who have little knowledge of the research that has been carried out by the statistics education community over the past several decades; building bridges between these two groups could be an important step for understanding what "inference" means in the context of big data. With very large data sets, for example, statistical significance is relatively easy to attain, so they must be regarded cautiously. In situations where statistical analysis is done on a population (e.g., census data or country-level demographic data), inference is often from the present to the future rather than from a sample to a population—and thus has much in common with considerations of modeling. The implications of these differences for informal statistical inference would be a fruitful topic for study.

Several of the differences among approaches to inference listed in Sect. 8.4 relate to how explicitly and quantitatively probability is described. The case studies in this chapter illustrate increasingly explicit notions of probability, from quite qualitative in primary school, to somewhat more quantitative in the middle/high school example (although not overtly tied to probabilistic theory), and to quite explicit and quantitative uses in the tertiary examples. Some questions around this topic that could profitably be addressed in research are as follows: How and when should probability topics be formally introduced to students? How does learning about probability relate to learning the concepts of informal statistical inference? How and when should they be merged?

Most of the research on inferential reasoning, both formal and informal, focuses on sample-to-population inference and group comparison: the realm of *t*-tests. There is much less work on inferences from time series (the case study included in this chapter is unusual) or inferences about correlations, such as might show up in a scatterplot. These are important directions for future work, as we need to understand how similar or different inferential reasoning processes are in these varied statistical contexts. How can students learn about inferential principles in one statistical context that they can easily transfer to other statistical situations?

Informal statistical inference has appeared in curriculum documents in some countries explicitly in secondary school (e.g., New Zealand Ministry of Education, 2016) or implicitly in primary school using words like "prediction" (e.g., Australian Curriculum and Reporting Authority, 2012). Bakker and Derry's (2011) work reminds us that informal inferential reasoning has the potential to create more coherence in the statistics curriculum as a constant through line that becomes more complex as students progress. There is not yet a curriculum document that develops informal inferential reasoning throughout school, although the papers reviewed in this chapter span the years from the first year of school (kindergarten) through the tertiary years, so there is clearly potential for a coherent sequence. Creating such a trajectory around inference could be a valuable joint enterprise for the statistics education community (see Ben-Zvi et al., Chap. 16 and Pfannkuch, Chap. 12 for further ideas on developing coherence in the statistics education curriculum).

References

Ainley, J., & Pratt, D. (2001). Introducing a special issue on constructing meanings from data. *Educational Studies in Mathematics, 45*(1), 1–8.

Aridor, K., & Ben-Zvi, D. (in press). Students' aggregate reasoning with covariation. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and trends in current statistics education research: International perspectives*. New York: Springer.

Australian Academy of Science. (2006). *Mathematics and statistics: Critical skills for Australia's future. The national strategic review of mathematical sciences research in Australia*. Melbourne: University of Melbourne. Retrieved from http://www.review.ms.unimelb.edu.au/FullReport2006.pdf

Australian Curriculum, Assessment and Reporting Authority. (2012). *Australian curriculum: Mathematics*. Melbourne: ACARA.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1–2), 5–26.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bakker, A., Kent, P., Derry, J., Noss, R., & Hoyles, C. (2008). Statistical inference at work: Statistical process control as an example. *Statistics Education Research Journal, 7*(2), 130–145.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods, 10*(4), 389–396.

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. International Association for Statistics Education: Voorburg, The Netherlands.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics, 45*(1), 35–65.

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM, 44*(7), 913–925.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–689). New York: Springer.

Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. J. (2013). Dynamic visualizations and the randomization test. *Technology Innovations in Statistics Education, 7*(2), 1–21. Retrieved from http://escholarship.org/uc/item/9dg6h7wb

Budgett, S., & Wild, C. J. (2014). Students' visual reasoning and the randomization test. In K. Makar, B. da Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: IASE.

Castro-Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2*(2), 98–113.

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295–323). Dordrecht, The Netherlands: Kluwer.

Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), Article 1. Retrieved from http://escholarship.org/uc/item/6hb3k0nz

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801–823.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*(1), 1–78.

Council for the Mathematical Sciences. (2004). *An international review of UK research in mathematics*. London: Author.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.
delMas, R. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79–95). Dordrecht, The Netherlands: Kluwer Academic Publishers.
delMas, R., Garfield, J., & Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education, 7*(3).
delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 25–58.
delMas, R., Garfield, J., & Zieffler, A. (2014). Using TinkerPlots™ to develop tertiary students' statistical thinking in a modeling-based introductory statistics class. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen, Mathematik und Stochastik lernen—Using tools for learning mathematics and statistics* (pp. 405–420). Wiesbaden, Germany: Springer Spektrum.
Dewey, J. (1910/1997). *How we think*. Mineola, NY: Dover.
Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*(1), 132–151.
English, L. D. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics, 81*(1), 15–30.
Falk, R., & Greenbaum, C. W. (1995). Significance tests die hard. *Theory and Psychology, 5*(1), 75–98.
Fielding-Wells, J. (2010). Linking problems, conclusions and evidence: Primary students' early experiences of planning statistical investigations. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*. International Statistical Institute: Voorburg, The Netherlands.
Fielding-Wells, J., & Makar, K. (2015). Inferring to a model: Using inquiry-based argumentation to challenge young children's expectations of equally likely outcomes. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 89–115). Minneapolis: Catalyst Press.
Frick, R. W. (1998). Interpreting statistical testing: Process and propensity, not population and random sampling. *Behavior Research Methods, Instruments, & Computers, 30*(3), 527–535.
Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching*. New York: Springer.
Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM, 44*(7), 883–898.
Garfield, J., Le, L., Zieffler, A., & Ben-Zvi, D. (2015). Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking. *Educational Studies in Mathematics, 88*(3), 327–342.
Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1), 87–108.
Groth, R. (2015). Working at the boundaries of mathematics education and statistics education communities of practice. *Journal for Research in Mathematics Education, 46*(1), 4–16.
Hacking, I. (2006). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge UK: Cambridge University Press.
Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*(3), 337–364.
Harlow, L. L. E., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum Associates.
Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education (A Joint ICMI/IASE Study)* (pp. 235–246). New York: Springer.
Hesterberg, T., Moore, D., Monaghan, S., Clipson, A., & Epstein, R. (2009). Bootstrap methods and permutation tests. In D. Moore, G. McCabe, & B. Craig (Eds.), *Introduction to the practice of statistics* (6th ed., pp. 16-1–16-60). New York, NY: Freeman.

Kahneman, D., Slovic, S., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Konold, C. (1994). Teaching probability through modeling real problems. *Mathematics Teacher, 87*(4), 232–235.

Konold, C. (2005). *Exploring data with TinkerPlots*. Emeryville, CA: Key Curriculum Press.

Konold, C. (2007). Designing a data analysis tool for learners. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 267–291). New York: Lawrence Erlbaum Associates.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning, 12*(3), 217–230.

Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration*. Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Langrall, C., Nisbet, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning, 13*(1), 47–67.

Lecoutre, M.-P., Durand, J.-L., & Cordier, J. (1990). A study of two biases in probabilistic judgments: Representativeness and equiprobability. In J.-P. Caverni, J.-M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases* (pp. 563–575). Amsterdam: Elsevier.

Lehrer, R., & Schauble, L. (2000). Modeling in mathematics and science. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Vol. 5, pp. 101–105). Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 635–679.

Lock, R. H., Lock, P. F., & Morgan, K. L. (2012). *Statistics: Unlocking the power of data*. Wiley Global Education.

Madden, S. R. (2011). Statistically, technologically, and contextually provocative tasks: Supporting teachers' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1), 109–131.

Makar, K. (2014). Young children's explorations of average through informal inferential reasoning. *Educational Studies in Mathematics, 86*(1), 61–78.

Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning, 18*(1), 1–24.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 152–173.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.

Meletiou-Mavrotheris, M., & Paparistodemou, E. (2015). Developing students' reasoning about samples and sampling in the context of informal inferences. *Educational Studies in Mathematics, 88*(3), 385–404.

Moore, D. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.

Mulligan, J., & Mitchelmore, M. (2013). Early awareness of mathematical pattern and structure. In L. English & J. Mulligan (Eds.), *Reconceptualizing early mathematics learning* (pp. 29–45). New York, NY: Springer.

New Zealand Ministry of Education. (2016). Elaborations on level 5: Statistics. *NZMaths.* Retrieved from http://nzmaths.co.nz/elaborations-level-five-statistics.

Nickerson, R. (2004). *Cognition and chance: The psychology of probabilistic reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, the Netherlands: IASE.

Pfannkuch, M., & Budgett, S. (2014). Constructing inferential concepts through bootstrap and randomization-test simulations: A case study. In K. Makar, B. da Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, the Netherlands: IASE.

Pfannkuch, M., Budgett, S., & Arnold, P. (2015). Experiment-to-Causation inference: Understanding causality in a probabilistic setting. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 95–128). Minneapolis: Catalyst Press.

Pfannkuch, M., Wild, C. J., & Regan, M. (2014). Students' difficulties in practicing computer-supported statistical inference: Some hypothetical generalizations from a study. In T. Wassong, D. Frischemeier, P. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen—Using tools for learning mathematics and statistics* (pp. 393–403). Wiesbaden, Germany: Springer Spektrum.

Pratt, D. (2011). *Re-connecting probability and reasoning about data in secondary school teaching*. Paper presented at the 58th World Statistics Conference, Dublin.

Pratt, D., Johnson-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal, 7*(2), 107–129.

Prodromou, T., & Pratt, D. (2006). The role of causality in the co-ordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal, 5*(2), 69–88.

Ramsey, F., & Schafer, D. (2013). *The statistical sleuth: A course in methods of data analysis* (3rd ed.). Boston: Cengage Learning.

Reaburn, R. (2014a). Introductory statistics course tertiary students' understanding of *p*-values. *Statistics Education Research Journal, 13*(1), 53–65.

Reaburn, R. (2014b). Students' understanding of confidence intervals. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: IASE.

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal, 7*(2), 5–19.

Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, the Netherlands: IASE.

Saldanha, L. A., & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257–270.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop et al. (Eds.), *International handbook of mathematics education* (pp. 205–237). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Stanovich, K. (2007). *How to think straight about psychology* (8th ed.). Upper Salle River, NJ: Allyn & Bacon.

Tintle, N., Rogers, A., Chance, B., Cobb, G., Rossman, A., Roy, S., et al. (2014). Quantitative evidence for the use of simulation and randomization in the introductory statistics course. In K. Makar, B. da Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, the Netherlands: IASE.

Tintle, N., Topliff, K., VanderStoep, J., Holmes, V. L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal, 11*(1), 21–40.

Tintle, N., VanderStoep, J., Holmes, V.-L., Quisenberry, B., & Swanson, T. (2011). Development and assessment of a preliminary randomization-based introductory statistics curriculum. *Journal of Statistics Education, 19*(1).

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician, 34*(1), 23–25.

Wagenmakers, E. (2007). A practical solution to the pervasive problems of *p*-values. *Psychonomic Bulletin and Review, 14*(5), 779–804.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–248.

Wild, C., Pfannkuch, M., Regan, M., & Parsonage, R. (2017). Accessible conceptions of statistical inference: Pulling ourselves up by the bootstraps. *International Statistical Review, 85*(1), 84–107.

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40–58.

Chapter 9
Statistics Learning Trajectories

Pip Arnold, Jere Confrey, Ryan Seth Jones, Hollylynne S. Lee, and Maxine Pfannkuch

Abstract Statistics curricula and pedagogy are changing rapidly in response to a growing body of research findings involving students' reasoning processes, technology capability, attention to underpinning conceptual infrastructure, and new ways of statistical practice. Because many of the statistical ideas being considered are currently not in the curriculum, many researchers in statistics education have investigated students' reasoning processes through the use of learning trajectories in conjunction with design-based research methods. In this chapter, we outline the characteristics of learning trajectories and exemplify how learning trajectories have been used in three case studies in statistics education. Commonalities and differences across the learning trajectories are discussed as well as recommendations for future research.

Keywords Statistics learning trajectory characteristics • Design-based research • Statistics learning goals • Statistics classroom-based observations • Conceptual pathways in statistics • Modifying statistical learning trajectories • Problem-based research in statistics education

---

P. Arnold
Cognition Education, Auckland, New Zealand
e-mail: parnold@cognitioneducation.com

J. Confrey • H.S. Lee
College of Education, North Carolina State University, Raleigh, NC, USA
e-mail: jere_confrey@ncsu.edu; hollylynne@ncsu.edu

R.S. Jones
College of Education, Middle Tennessee State University, Murfreesboro, TN, USA
e-mail: Ryan.jones@mtsu.edu

M. Pfannkuch (✉)
Department of Statistics, The University of Auckland, Auckland, New Zealand
e-mail: m.pfannkuch@auckland.ac.nz

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_9

9.1 Introduction

Over the last 20 years, learning trajectories (LTs) have gained prominence in statistics education research. Part of the prominence may be due to a general trend toward a participationist research and design paradigm in education (Sfard, 2005). In this paradigm, there is an emphasis on understanding the teaching and learning process as it develops in actual classrooms with researchers positioning themselves as collaborating with teachers rather than studying them—"there is a remarkable blurring of the boundaries between communities of researchers and practitioners" (Sfard, 2005, p. 401). The trend in education research toward studies with ecological validity and a participationist paradigm may have set the scene for statistics education researchers to use LTs particularly as many of them were searching for new ways to approach statistical learning.

Traditionally statistics has been taught as a series of techniques to handle and display data with little regard for students' reasoning processes and the building up of conceptual infrastructure across the grade levels. With attention now focusing on students' reasoning from data and on conceptual understanding of statistics, researchers have found that the conceptual underpinnings are not only difficult to grasp but also difficult to elucidate (cf. Chap. 8). Therefore, to explicate the conceptual foundations for and across statistical topics, it has been necessary to build new LTs within and across grade levels for research and teaching purposes. Furthermore, research in statistics education is challenging traditional curricula and pedagogy with respect to the content and the lack of attention to conceptual pathways and to research findings. This challenge is coming from researchers who are concerned about problems in students' reasoning processes and the links these problems have with instructional processes. These researchers invented innovative LTs because they were attempting to scaffold new conceptual understandings in students that were not present in current curricula. They used LTs to explore and document students' thinking as they engaged with new approaches to statistics (e.g., Bakker, 2004; Makar, Bakker, & Ben-Zvi, 2011). Hence, research and curriculum development and task design and students' thinking are both strongly connected within LTs (cf. Clements & Sarama, 2004). By following the development of students' thinking as they engage in a sequence of instructional tasks, new findings and gaps in students' thinking can emerge, which can result in new research and curricular paths for learning (see Bakker & Gravemeijer, 2004).

In Sect. 9.2 we elaborate on key characteristics of LTs and then illustrate in Sect. 9.3 the use of LTs in research with three case studies. Finally, we reflect on the case studies and discuss implications and recommendations for future research.

9.2 Characterizing Learning Trajectories

In recognition that LTs were being interpreted and applied in a variety of ways within research, Clements and Sarama (2004, p. 83) stated:

> We conceptualize learning trajectories as descriptions of children's thinking and learning in a specific mathematical domain and a related, conjectured route through a set of instructional tasks designed to engender those mental processes or actions hypothesized to move children through a development progression of levels of thinking, created with the intent of supporting children's achievement of specific goals in that domain.

A similar conceptualization of LTs is held among statistics education researchers for the statistical domains. However, to understand the characteristics underpinning LTs, we need to return to their origin.

LTs were originally conceived as hypothetical learning trajectories in the seminal work of Simon (1995) who described from a constructivist perspective how teachers could conceptualize and enact the learning process within their classrooms. He perceived the LT as hypothetical because it was based on a teacher's prediction of the learning process before it was implemented. During implementation the LT would be constantly updated in response to observations on students' interactions and reasoning processes. Because the term LT is now commonly used in the literature, we use it to describe the predicted trajectories and the updated trajectories. Other researchers (e.g., Lehrer, Kim, Ayers, & Wilson, 2014) prefer to use the term learning progressions to reflect a more open process. Although we refer to researchers using LTs, in practice teachers and researchers often collaborate on designing and studying LTs, and teachers in their own classrooms also enact the LT teaching cycle.

The LT (see Fig. 9.1) involves defining a learning goal, considering possible learning activities and the types of student thinking and understanding they might evoke, and the hypothetical learning process (Simon, 1995). To produce a LT, a learning goal is initially defined, and then a hypothesis is formed about a particular group of students' understanding within that topic domain (Fig. 9.1(1)). The hypothesis is based on information from a wide range of sources and experiences, for example, current students' experiences in a related area, the experiences of a similar group of

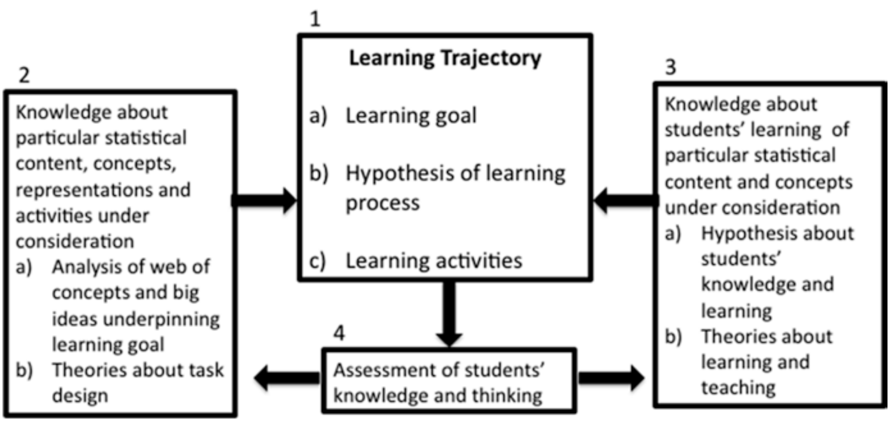


Fig. 9.1 The learning trajectory and sources drawn upon (based on Simon, 1995, p. 137)

students, information about prior knowledge that has come to light from pretesting, and data and information from the research literature (Fig. 9.1(3 and 4)). Another dimension in the creation of LTs is the undertaking of an analysis of the web of concepts including the big ideas (Ben-Zvi & Garfield, 2004) that may need to be addressed in reaching the learning goal (Fig. 9.1(2a)). For example, if the learning goal is for students to learn how to reason from distributions, then an analysis of the concepts and big ideas underpinning distributions (e.g., data, center, variability) needs to be undertaken in cognizance of future LTs that may address concepts and ideas that cannot be incorporated into the current trajectory (e.g., inference).

Based on the researchers' hypothesis of students' knowledge, skills, and possible thought processes and an analysis of the concepts and big ideas underpinning the main goal, potential learning activities and the types of thinking and learning these activities might provoke are considered. Researchers' theories about statistics teaching and learning (Fig. 9.1(3b)), their knowledge of learning in the statistics context, and their knowledge of statistics activities and representations (Fig. 9.1(2)) all intersect and come into play when considering possible learning activities (Simon, 1995). Statistical tools as mediators in the learning process need to be evaluated for inclusion in learning activities, while attention to classroom discourse and how it could be used to elicit and scaffold students' understanding is another important consideration. Other influences also impinge researchers' plans for learning activities, besides age-related development issues, such as cultural factors (Clements & Sarama, 2004), and researchers' beliefs and interests including those of the teachers that they may be collaborating with (see Chap. 10).

The learning activities can also draw on research about task design, an area of research that has only recently come to the forefront (see Watson & Ohtani, 2015). Task design is considered important because the content of the tasks affects students' learning and the nature of the learning (see Chap. 16). For research about learning, the tasks given to students have a major influence on the resultant findings about their conceptions and capabilities. Principles for designing tasks have been elucidated by Lesh and Doerr (2003) for model-eliciting activities such as personal meaningfulness to the student and the ability to generalize from the model constructed. Ainley, Pratt, and Hansen (2006) also emphasize the importance of attending to purpose and utility when designing tasks. LTs often incorporate implicit task design principles into the learning activities that are developed, suggesting more consideration is needed in this area (Fig. 9.1(2b)).

The hypothesized learning process is "a prediction of how students' thinking and understanding will evolve in the context of the learning activities" (Simon, 1995, p. 136). This is a best guess at what will happen. There is no suggestion that the instruction sequence is the only or best path for teaching and learning, only that it is *one possible route* (Clements & Sarama, 2004). A LT can also be thought of as a description of the set of intermediate behaviors (including both landmarks and obstacles) that are likely to emerge, as students progress from naïve preconceptions toward more sophisticated understandings of a target concept (Confrey, 2006). The hypothetical learning process is continually modified. This is a result of the researchers developing a broader understanding of students' conceptions in the area through a process of reflection based on interactions with and observations of students. The researchers' thinking is modified as they make sense of what is happening in the classroom. Reflection, based on assessment of students' thinking, leads to constant adjustment and fine-tuning of the LT, the goal, the activities, and the hypothetical learning process (Simon, 1995).

The assessment of students' thinking to inform modifications to the LT (Fig. 9.1(4)) can be investigated in a variety of ways such as individual written diagnostic tests, task-based individual or group interviews requiring thematic qualitative analyses, and analyses of classroom discussion and interaction. An interesting example of addressing the problem of how to analyze classroom interaction data is found in the work of Dierdorp, Bakker, Eijkelhof, and van Maanen (2011). To determine how well conjectures about students' learning matched up with the observed learning, they used a data analysis matrix and a summary coding system for transcripts from classroom interactions in order to gain insight into how their LT supported students' inferential reasoning. More work is needed in this area to provide better evidence in research papers about how a LT supports or does not support students' learning with respect to the learning goal.

The LT systemizes and extends what good teachers do, with the difference being that within a research context, it is a deliberate act: the researchers are actively and consciously planning, reflecting, and recording actions and thoughts. As a LT is being trialed through several iterations on groups of students, the goal of the researchers is to deliberate on the observed student development together with the instructional sequence and form a localized theory of instruction (Gravemeijer, 2004). It is localized because the theory may only pertain to the group of students on which the instructional tasks were implemented, but other researchers may be able to take the theory as a framework for developing LTs for their particular group of students. Bakker and van Eerde (2015) explain that similar patterns of students' thinking can emerge across different classrooms and teaching experiments resulting in a more general theory of instruction of how a topic can be taught.

In education research, the use of LTs as a research instrument is often associated with design-based research (DBR) methodology (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Confrey & Lachance, 2000; Gravemeijer & Cobb, 2006; Prediger, Gravemeijer, & Confrey, 2015). DBR is characterized as research where students' development and progression are analyzed using deliberately designed learning activities with the aim of testing or developing theory (Bakker & van Eerde, 2015). The aims of DBR in which a new type of learning is engineered can be manifold: explanatory and advisory "to give theoretical insights into how particular ways of teaching and learning can be promoted" or predictive to state that "under condition X using educational approach Y, students are likely to learn Z" (Bakker & van Eerde, 2015, p. 431). Another characteristic of DBR is its iterative nature where cycles of preparation and design, teaching experiment, and retrospective analysis are conducted. During the teaching sequence, researchers can ascertain how the learning occurs in actual practice and through reflecting critically can then adjust or modify the plan for the next lesson. Typically these are small changes from lesson to lesson. After the teaching sequence is implemented, larger-scale modifications can be made to the LT. DBR has recently undergone further development (see Design-Based Implementation Research, 2016). Hence, DBR methodology forms a natural partnership with LT research. Mixed methods research methodology can also be used in conjunction with LTs.

9.3 Three Case Studies of Learning Trajectories

The statistics education community has produced a number of studies that contribute to the knowledge base on LTs (Franklin et al., 2007; Lehrer et al., 2014; Rubin, Bruce, & Tenney, 1990). Research suggests that statistical concepts should be integrated into inquiry activities and that how students think about statistical concepts evolves as students grow in encountering accessible forms of variability (Garfield & Ben-Zvi, 2007; Konold & Pollatsek, 2002; Pfannkuch & Wild, 2004) that create a need for the concepts (Confrey, 1991).

In this section, with references to Fig. 9.1, we illustrate how LTs can be used in statistics education research. In the first case study, Jere Confrey and Ryan Seth Jones illustrate strategies to represent hypothesized construct maps to help teachers and students trace the growth of students' thinking about variability. Pip Arnold, in the second study, has the learning goal of making a judgment or an inference when comparing two box plots, and she exemplifies how students were scaffolded, using a hypothetical learning process, toward that goal. In the third study, Hollylynne Lee, in collaboration with Helen Doerr, designed a LT to advance teachers toward an understanding of repeated sampling for inference. All these studies used DBR. At the heart of these case studies is the big idea of variation, from the need to invent a statistic to describe the variation observed to the need to take variation into account when making an inference.

9.3.1 Case Study 1: Two Preparatory Learning Trajectories for Sixth-Grade Students toward Inventing a Statistic for Variability

9.3.1.1 Introduction

The first case study addresses students' introduction to the concept of variability, a topic studied by numerous scholars (e.g., Ben-Zvi, 2004; Garfield & Ben-Zvi, 2005; Konold & Pollatsek, 2002; Lehrer et al., 2014; Makar & Confrey, 2005; Wild & Pfannkuch, 1999). Confrey and Jones chose to approach the topic using a learning map organized around big ideas, which were broken down further into constructs with underlying LTs (Confrey, 2015). These LTs accurately characterize typical responses from students in increasing levels of sophistication. The map is used for two primary purposes: to provide professional development opportunities for teachers and to develop diagnostic assessments to gauge student progress.

9.3.1.2 The Learning Goals and the Designed Learning Process

Confrey and Jones started with the learning goals from the Common Core State Standards for Mathematics (Common Core State Standards Initiative, 2010) in the United States for sixth-grade (age 11) statistics. Through analyzing the web of concepts and big ideas underpinning the learning goals (cf. Fig. 9.1(2a)), they designed a learning map that was hierarchically organized around nine big ideas identified by Confrey. The big ideas were subdivided into one to five relational learning clusters, which were made of sets of mutually supporting constructs. Each construct was described with a corresponding learning trajectory consisting of an ordered set of indicators of increasing sophistication. These reflect the likely student behaviors and thinking that would emerge as they progressed through instruction (see Table 9.1 for the first two constructs). In statistics one big idea was "display data and use statistics to measure center and variations in distributions." This big idea was divided into three relational learning clusters: (1) displaying univariate data, (2) measuring data with statistics, and (3) displaying bivariate data. Each learning cluster was divided further into a set of connected constructs. The constructs for displaying univariate data were (1) gathering data and describing variability, (2) displaying data in novel and traditional ways, (3) comparing different displays of the same data, and (4) shape of univariate data.

The LTs were based on a synthesis of literature from statistics education research and previous iterations of the learning trajectory. For example, prior to this study, many of the behaviors and thinking about variability were articulated in the related learning cluster on modeling. However, after the foundational role of this thinking was observed in their studies for making sense of data displays and statistics, they restructured the map to include these ideas in the data display cluster. In each iteration of the LT, patterns in student thinking are reinforced, but nuanced variations or even new ways of thinking emerge and are added into the LTs.

The overarching learning goal of the trajectory for displaying univariate data was to support students to develop a conception of variability that was represented in various data shapes created by displaying data and to lay the groundwork for needing a measure of variability in later trajectories (Konold & Pollatsek, 2002; Lehrer & Kim, 2009; Petrosino, Lehrer, & Schauble, 2003). The goal was influenced by Confrey and Jones' theories about learning (cf. Fig. 9.1(3b)) the key elements of which include the role of invention and of transformation (accommodation in Piagetian terms). Another key element of their approach was to foster discourse among the students, so they could learn from each other's ideas and contributions. Teachers play a central role in bringing forth this thinking and building classroom norms valuing articulation and sharing of ideas. Their belief is that the LTs should also communicate the kinds of student statistical thinking teachers should attend to and how they fit together into trajectories of increasingly sophisticated thinking. Thinking about variability, displaying one's data, and comparing those displays prepare the ground for a discussion of data shape and statistics (Lehrer, Kim, & Jones, 2011; Petrosino et al., 2003; Schwartz & Martin, 2004). Only after students have productively struggled with these ideas are they ready to invent statistics and learn conventional definitions.

9.3.1.3 The Learning Activities and the Observed Learning Process

Confrey and Jones developed instructional materials by drawing on prior work by Lehrer (2016) and Confrey (2002), and they made use of *TinkerPlots* (Konold & Miller, 2005) and *Data Games* (Finzer, Konold, & Erickson, 2012) for data exploration and display. Hence, the learning activities were based on their knowledge of teaching strategies and resources for statistics, their knowledge of how students might learn about univariate data displays, and their understanding of the current knowledge of the students who would be in their study (cf. Fig. 9.1(2 and 3)). Diagnostic assessment to gauge student learning was also coordinated with the LTs (cf. Fig. 9.1(4)).

The following case involved 15 sixth graders (age 11) who met for 3 hours per day in a classroom on their research site for 1 week. The purpose of the study was twofold: (1) to confirm or modify the LT and (2) to collect samples of student work for professional development purposes. Thus, the research question under investigation was: What patterns of behavior, forms of representation, and ways of talking are in evidence among students when introduced to the ideas of multiple sources of variability and displaying univariate data, and how might these patterns be represented so that they are intelligible and useful to teachers?

The case study provides an image of student learning and how this learning is represented in the two constructs in Table 9.1. Throughout the description of student activity, the relevant levels are referenced within that construct. Note that in this description, Confrey and Jones are assessing students' knowledge and thinking (cf. Fig. 9.1(4)) in order to inform them whether the observed patterns of behavior are consistent with the indicators listed in Table 9.1.

Gathering Data and Describing Variability

To engage students with the problem of creating variable data (the first two levels of this construct), they asked students to consider three different questions: What is the circumference of the fountain in our courtyard?, How many M&Ms. are in one individually wrapped package?, and What is the circumference of a middle schooler's head? To highlight the challenge of variability, they left the data collection strategies open-ended and provided crude measurement tools, such as string and rulers. Under these circumstances, students produced significant measurement error.

Student measurement mistakes, though, were a resource for them to make sense of the various sources of variability in the data. To elicit a conversation about sources of variability in the data (level 4 in Construct 1, Table 9.1), the teacher posted unordered lists of the students' data and asked them in a whole class conversation "what do you notice when you look at all this variability?" They also discussed relative magnitude of the different sources they identified (level 5) by asking questions such as "what caused the variability in the different types of data?" Table 9.2 provides short examples of the kinds of student comments that are common to this discussion.

Table 9.1 LT indicators for two constructs

Constructs
Construct 1: Gathering data and describing variability
1. Recognizes target phenomenon and asks questions about it
2. Creates and uses data as information to answer a question
3. Describes that some questions have uncertain answers because of variability in the data
4. Identifies sources of variability in data
5. Estimates magnitude of different sources of variability
6. Categorizes sources of variability (measurement error, natural variability, production error)
7. Anticipates variability in data across different samples
8. Describes or predicts how a change in process affects variability
Construct 2: Displaying data in novel and traditional ways
1. Displays data without reference to investigation
2. Shows basic familiarity with bar graphs, pie charts, and dot plots
3. Identifies or creates titles, labels, or keys
4. Orders data from least to greatest without distinguishing scale from data
5. Stacks individual values or within groups, intervals, or bins
6. Scales using equal intervals
7. Creates dot plots and bar graphs, knows the distinction between categorical and continuous data, and explains choices about scale, order, and grouping
8. Creates histogram and circle graph, explains choices about intervals, and provides either count or percent within each interval
9. Makes and justifies choices about displaying data in traditional and novel ways

Table 9.2 Key concepts and student comments about variability

Data context	Student comments
Individual measurements of the same object (in our case, the fountain in the courtyard)	"…with the fountain it's like whoa! What happened here? I see 461 and I see like 2010!" "people could have made mistakes when they measured" "people might leave a gap when they flip over the ruler"
Number of M&Ms. in different bags	"the factory probably didn't measure out the exact amount of M&Ms" "we also could have counted wrong" "It's possible to miscount 10, but it's not extremely likely"
Individual head circumferences	"…it's kind of common that everyone got different head circumferences, because not everyone's head is the same size" "the variation might be the result of different sized heads, but they also might be the result of mistakes people made"

Students observed that the variability in the fountain data was a fundamentally different kind of variability than the other two types, and they drew on their data creation experiences to generate theories about the kinds of errors that likely produced the variability. Students then shifted from describing data as measurements to calling it “opinions,” indicating their feeling that there was so much variability in it that it was not “scientific” enough to be called data. This conversation ended with the teacher asking, “What would the data look like if a class of students similar to us measured the same fountain?” This question was asked to evoke early ideas about sample-to-sample variability (level 7). Students quickly responded that the data would look “similar to ours”, that their data would have “the same kinds of chaos as our data,” and that it would “have a similar median and mean, but the numbers will be different.”

These themes ran throughout the rest of the activities. For example, they were the driving motivation for remeasuring the fountain more precisely to see how a change in process affects variability (level 8) and creating paper hats using their measurements to estimate the extent to which measurement error contributed to the variability in the head circumference data (level 5).

Displaying Data in Novel and Traditional Ways

Confrey and Jones provided opportunities for students to invent strategies for displaying their fountain data in a way that helped them think about the true length and the variability in the measurements. Students revealed a variety of strategies for displaying their data, many of which had been documented by other researchers (e.g., Lehrer & Schauble, 2002). Here two of the four displays that students invented are presented to illustrate the ways student thinking corresponds to the learning map and to show how student thinking developed as they invented and revised their displays.

Group 1 produced a dot plot without distinguishing between the data and the scale (Fig. 9.2). They explained that they wanted a display that clearly displayed every measurement observed and how often each measurement was observed. They made the decision to order the data from least to greatest (level 4), without representing gaps, but with stacking of identical values (level 5).

Group 2 created a histogram (that they referred to as a box plot) with 100 cm intervals (Fig. 9.3). Similar to group 1, they ordered the data from least to greatest (level 4), but, in contrast, they grouped and stacked all values within a 100 cm interval (level 5) and created an interval scale (level 6). These choices created a very different representation of the data, which provided a context to discuss the tradeoffs between the two.

As they created these invented displays, the students sometimes showed evidence of thinking at the lower three levels of the construct as they sometimes considered decisions without referencing the question about the fountain circumference, referred to approximate notions of conventional displays, and created titles and labels. However, the most significant intellectual work for students came when they had to consider decisions about order, grouping, and scale (levels 4–6). For exam-

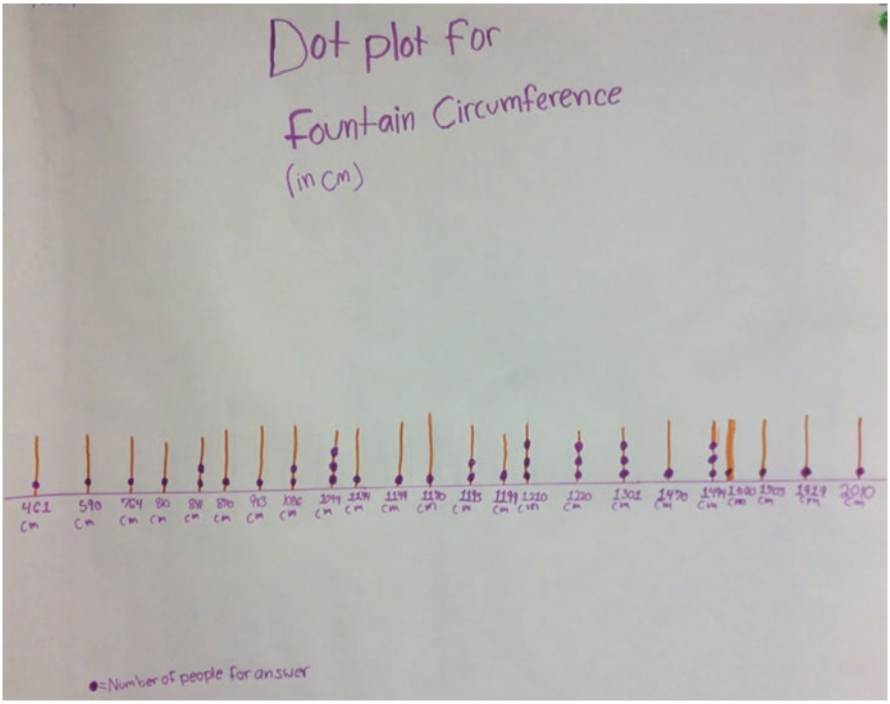


Fig. 9.2 Group 1 data display

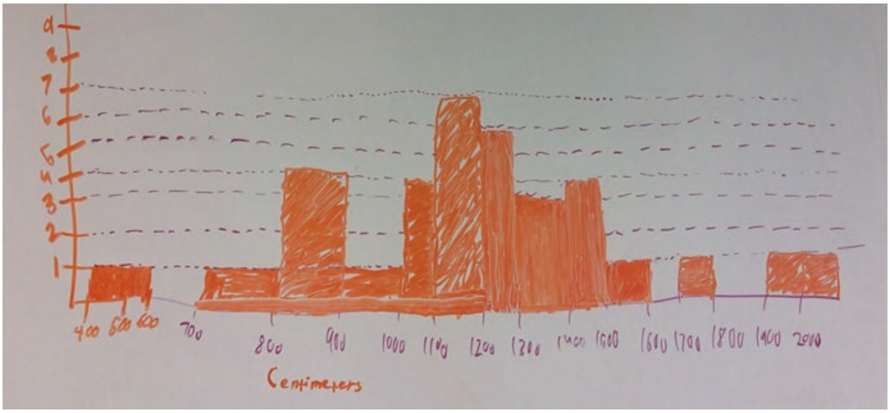


Fig. 9.3 Group 2 data display

ple, the students had to decide if the display scale needed to include values that were not observed. Their early decisions about data displays were not driven strictly by convention, but more by their desire to make sense of variability and communicate meaning to their peers. Only after wrestling with these issues were data display conventions (levels 7 and 8) introduced, so the conventions could be rooted in student ideas and displays. When given the opportunities to build their ideas about statistical thinking from accessible forms of variability, students often demonstrate the behaviors, strategies, and thinking described in these LTs.

9.3.1.4 Discussion and Future Recommendations

This case illustrates the potential value of designing a learning map based on an analysis of the big ideas and the web of concepts that need to be included in the LT for supporting teachers to understand student thinking. It also illustrates the need for several iterations of LTs and reflection and analysis on students' responses in the development of that map. By explicating in detail indicators of likely student thinking as they progress through the LTs, this research across multiple settings is at the stage of developing a dynamic representation of student thinking that can serve as an orienting framework for curriculum and assessment design. A product of the research for teaching is the learning map and its LTs including resource material for teaching and student work for teacher professional development which can make patterns of student thinking intelligible to teachers.

The advantage of the approach outlined in this case is that the LTs for data, variability, and statistics are related to LTs that Confrey and her team have developed and refined across all big ideas for middle grade mathematics. This provides teachers a comprehensive resource to have access to syntheses of learning trajectories. In addition, the map makes it possible to study what the effects of an overall approach informed by LTs would be as students accumulate experience with the map. Too often LT studies are difficult to continue across grades as students switch teachers and classes. In this way, research can contribute to the building up of infrastructure for supporting the long-term development of statistical concepts, a facet that is lacking in current curricula.

9.3.2 Case Study 2: Preparing Ninth-Grade Students to Make the Call—Learning How to Make a Judgment When Comparing Two Box Plots

9.3.2.1 Introduction

The second case study illustrates a LT which started with a well-defined learning goal but required thought about the underpinning concepts that students needed to experience. Because the learning goal was new to the curriculum and resources did not exist, Arnold and a research team of two statisticians and nine teachers collaboratively worked on inventing language to describe the statistical ideas and designing learning strategies and resources. The challenge in this case study was to develop a set of structured learning experiences that would enable grade 9 (age 14) students to "discover" collectively the criteria for "making the call"—making a judgment when comparing two box plots.

9.3.2.2 The Learning Goals and the Designed Learning Process

The learning goal arose from a study on the reasoning processes of students in a grade-9 class. The students were learning how to make an inference when comparing two box plots and were making the call based on a variety of criteria (Pfannkuch, 2007). From the student responses, it was clear there was no agreed understanding between the teacher and her students as to what constituted support for an inference. Furthermore, the investigative question that the students were exploring was about the populations, but the students' reasoning was based on describing the sample statistics. In New Zealand, the curriculum (Ministry of Education, 2007) and subsequent national assessment required students to make informal inferences (see Chap. 8) about populations from samples for comparative situations. This created the problematic situation. Hence, a developmental pathway was proposed for comparative situations from grade 9 to grade 12 for justifying how to make a call or make a decision about whether condition A tends to have bigger values than condition B back in the populations (Wild, Pfannkuch, Regan, & Horton, 2011). The problem for this study was how to create a LT to enable students to understand the rationale and concepts underpinning making the call using the rule as outlined in Fig. 9.4.

In cognizance of the research literature and an analysis of the web of concepts (cf. Fig. 9.1(2a)) needed for making the call, the research team determined that enabling students to make the call depended on building their understanding of a network of underlying interrelated concepts, the key concepts identified being sample, population, and sampling variability. They considered sampling variability reasoning to be at the core of statistical practice but noted it had only recently received attention in school curricula and instruction. Typically, students reach the final years of high school, where they are explicitly introduced to notions such as basic statistical inference from confidence intervals, without fundamental knowledge or experiences of sampling behavior. Despite the importance of considering variation in statistics, researchers have only in the last two decades begun to document stu-

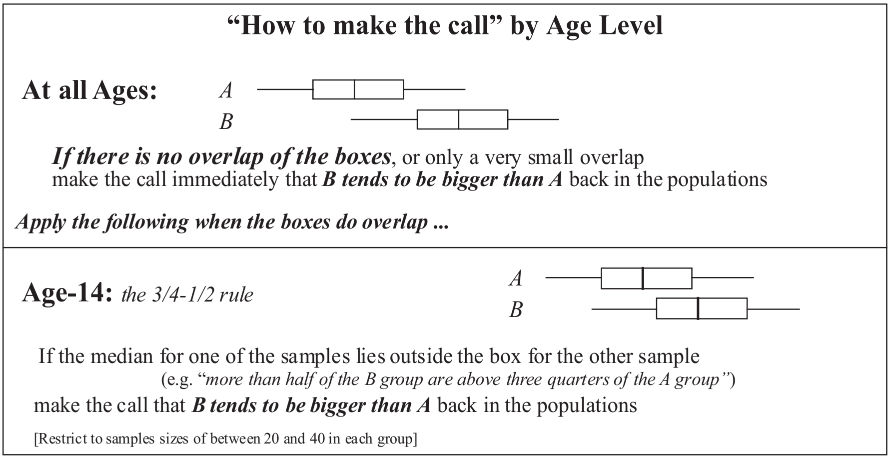


Fig. 9.4 How to make the call at ninth grade (age 14) (cf. Wild et al., 2011, p. 260)

dents' conceptions of variability. Therefore, a carefully structured set of learning experiences to support the LT was required if students were to understand and appropriate the sampling variability reasoning underpinning statistical inference. As Garfield and Ben-Zvi (2007) stated in relation to distribution, center, and variability, students "need help in developing an understanding of what these concepts actually mean and how to reason about them in an integrated way" (p. 386).

9.3.2.3 The Learning Activities and the Observed Learning Process

This case study reports on one class, although the research was undertaken with a number of classes (see Arnold, Pfannkuch, Wild, Regan, & Budgett, 2011). The planning and preparation phase involved trialing potential learning activities with the research team and making continuous changes to how the development of the three key concepts could be approached. Changes to the LT were also made when implemented in the classroom. The research question was: How can grade 9 students be facilitated to consistently and coherently make a statistical inference?

As already signaled, the three key concepts of population, sample, and sampling variability were important to support the LT for making the call. Specific learning materials and activities were created to support the development of these concepts and to support the LT (Table 9.3), which comprised 15 lessons. Some activities were deliberately planned and developed from the outset with the LT in mind, and some activities were developed as part of the ongoing reflection on the LT throughout the implementation in this class. In the description that follows are some vignettes of the learning experiences including examples of how and why the LT

Table 9.3 LT for the development of key concepts when comparing two box plots

LT for learning how to make a judgment when comparing two box plots ($n \approx 30$)
1: Sampling data from a population
• Identify population
• Pose and critique investigative questions about the population
• Recognize the need to use a sample to answer a question about the population
• Acknowledge that samples from the same population for the same variable provide similar information
• Appreciate that inferences about a population can be made from a sample
2: Sampling variability
• Recognize sampling variability in the center across multiple samples of the same size from the same population
• Appreciate sampling variability in the center and the extent of the overlap when comparing two box plots across multiple samples of the same size from two different populations
3: Developing criteria for making the call
• Recognize the salient features to focus on when comparing two groups (e.g., center, shift, overlap) • Acknowledge that there are two situations for making the call when comparing two groups

was modified in response to the research team's observations during the preparation stage and the collaboration of Arnold and the teacher in the classroom during the implementation stage.

Population and the "Population" Bags

As the "population" of Karekare College students (a fictitious college) was going to be used extensively throughout the teaching implementation, it was important that students in the class became familiar with the data that was available. The population of Karekare College students was represented using a plastic bag filled with data cards (see Fig. 9.5). Each data card represented 1 student and contained 13 different variables relating to the student. To develop familiarity with the data, students had to work out what the different variables were on the data cards.

During subsequent lessons, whenever the teacher referred to Karekare College, she nearly always showed the population bag (see Fig. 9.5), indicating that she was referring to the whole population, not just the data cards that the students had selected. The ability to keep reminding students that they were making an inference about the population by holding up the bag was an addition to the LT by the teacher, which was regarded by her and other teachers as an important facet in aiding students' statistical reasoning processes. Giving students an image of the population was an issue that was extensively debated by the research team, because the Karekare College data had been randomly selected from a large CensusAtSchool New Zealand (2003) database and hence could be considered a sample, but then the database was also a sample itself. By considering students' understanding of these

Fig. 9.5 Karekare College data cards and the population bag

issues and the fact that they were novices (cf. Fig. 9.1(3)), the research team decided to view the Karekare College students as the population. Although the population bag provided a good visualization of the population, it was insufficient, as a posttest revealed students did not have images for or contextual knowledge about population distributions. Hence, the assessment (cf. Fig. 9.1(4)) led to the creation of an additional LT.

Developing the Idea of Using a Sample

Having established the population and the variables for which data were available, the students posed a variety of investigative questions. The teacher and Arnold together identified which of the variables would be used for the activity where the concept of sample was first addressed. From the different investigative questions that the students posed, one was selected to be explored further. The students were to answer the question: "What are typical popliteal[1] lengths of students at Karekare College?" The teacher, as part of the planned LT, asked them how they might go about answering this question, to which they ultimately replied that they would be "putting [the data] in a graph." There was then some discussion and the students, working in small groups each with their own population bag, started to graph all (616 students) of the student data, using the data cards and a pre-prepared grid. After about 10 minutes, some general discussion started about "students" not all fitting onto the grid. A student said, "I'm not going to organize the whole college into this," at which point the teacher asked, "Is there a better way than looking at the whole lot?" The ensuing discussion and action resulted in students continuing until they had filled up their group's grid or felt that the shape of the graph was not changing despite adding more data cards, i.e., they did not use the whole population, just part of it. The teacher allowed the idea of using a sample, rather than the whole population to answer the question, to come from the students—she did not say to her class at the start, "Take a sample and use this to answer the investigative question." From the observed responses of the students, it was felt that the students were developing the idea that a sample could tell them something about the population. This observation was reinforced when comparing pre- and posttest student assessment responses as in the posttest students specifically referred to the population of interest in their investigative questions and in their conclusions.

Sampling Variability

Sampling variability was explored in a number of ways. In the lesson described previously where sampling was first introduced, the students had created their graphs using the actual data cards, which provided a strong visual display. The

[1] The popliteal length is a measurement taken on the back of the leg from behind the knee to the floor when a student is seated.

teacher gave students time to walk around the class and see how their graph compared with other graphs in the class. The students looked at features that were similar and features that were different. All groups gave an indication of where they felt the middle of their popliteal length data was, and across the class the set of middle popliteal-lengths for the different groups lay within a 3–4 cm band. The students were able to see that the middle popliteal length was similar even though the samples were different.

Sampling variability was a focus again in a later lesson about making the call when students were looking at the patterns across different samples with respect to two variables: student heights disaggregated by gender and time taken to get to school disaggregated by mode of transportation. These two examples were deliberately chosen for the LT as they captured very clearly the two situations described in Fig. 9.4. Note that students were observing box plots, with only the box part drawn, a modification made to the LT when trialed with the research team in order to focus student attention on the salient features for making the call (see Fig. 9.6 and Arnold et al., 2011).

Making the Call

When students were looking for patterns across the sets of graphs, Arnold and the teacher realized that additional prompts were required because information about the shift and the position of medians was not forthcoming. According to Bodemer, Ploetzner, Feuerlein, and Spada (2004), leaving students to generate hypotheses about relationships on their own is very hard, and they may not pay attention to salient features. Bodemer et al. (2004) suggest that learners' interactions with learning materials should be structured so that hypotheses are formulated only on one relevant aspect of the visualization at a time, and therefore in a modification to the LT, the students were guided to first focus on the distributional shift and then on which median was bigger.

After students had sorted their samples for each question, the teacher and class reflected on the process. They described and abstracted the patterns and criteria for making a call about what was happening back in the two populations. This allowed students an opportunity to extract relevant principles (Bakker & Gravemeijer, 2004). The students noticed that in the samples for heights, the boxes were close together, whereas in the samples for time taken to get to school, the boxes were apart (Fig. 9.6). They named these two situations about the relative location of the boxes Situation 1 and Situation 2, respectively.

In the following excerpt, they explore the differences between the two situations (see Fig. 9.6):

Teacher: So in our first situation we've got the boxes. They're all overlapping; some of them are going this way and some of them are going the other way. The medians are very close together, and the medians are also within the overlap of the boxes. In the second situation, how is it different? What's different about the

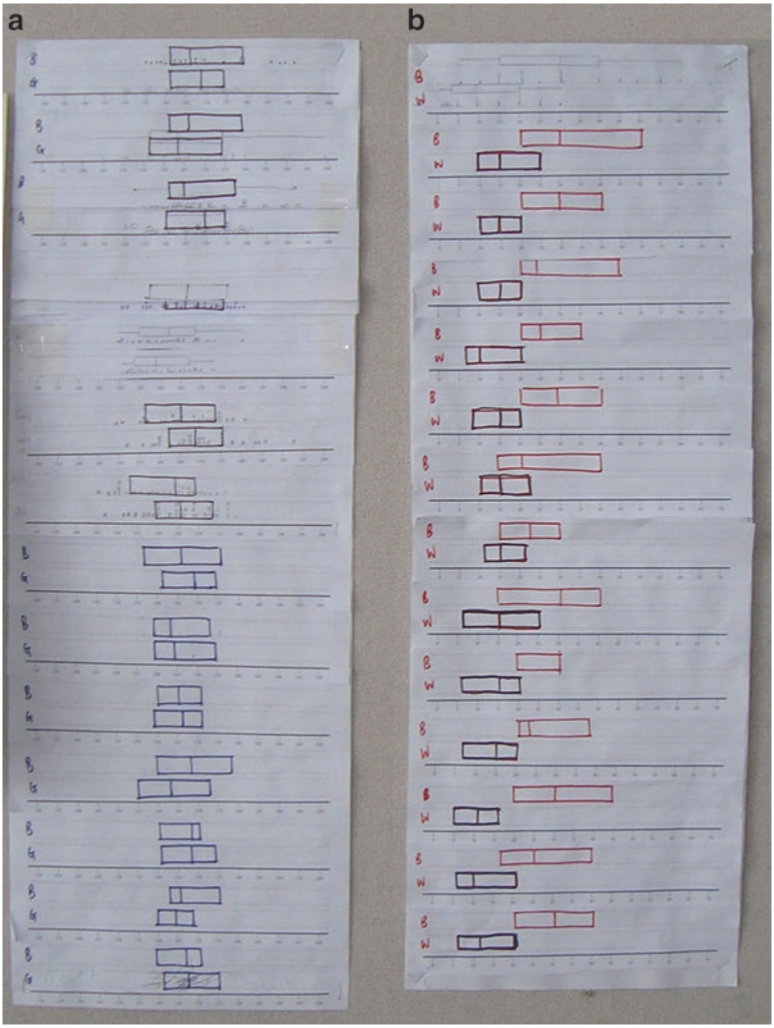


Fig. 9.6 Box plots of two situations: (**a**) samples comparing heights of girls and boys (on the *left*) and (**b**) samples comparing time taken to get to school by bus and walking (on the *right*)

overlap here? Is there no difference between the overlap on these boxes and these boxes?

Student: They're not overlapped so much.

Teacher: They're not overlapped so much. No, they're not. Okay, do they all overlap?

Student: No.

Teacher: No, so when they do have an overlap, they don't overlap much and otherwise they don't overlap at all. What can you tell us about the medians in this one?

Student: They're not overlapped.
Teacher: They're not in the overlap.

Visually and verbally, the students and teacher described differences in the two situations in terms of shift, overlap, and location of the medians. The students and teacher started to develop the criteria and language for making or not making a call. Collectively they spontaneously used hand gestures to describe the two situations, *close* (Figs. 9.6a and 9.7a) and *apart* (Figs. 9.6b and 9.7b), with vibrations to show the effect of sampling variability. Gestures according to Radford (2009) are a precursor to verbal conceptualization. The use of these gestures and the naming of the two situations, as Situations 1 and 2, by the teacher and students were built into the LT in subsequent classroom implementations.

The students also noticed that in Situation 2, there were consistent messages from the samples about the relative location of the two medians to one another back in the populations, allowing them to determine the larger of the two population medians, i.e., the median time to school by bus was always longer than the median time to school by walking. This was not the case in Situation 1. The students noted that sometimes the boys' median height was higher than the girls' median height and sometimes it was the other way around. Through recognizing and reasoning from the patterns in the two situations, they "discovered" collectively the criteria for making a call when two box plots are compared and the boxes overlap (age 14, Fig. 9.4) and do not overlap (at all ages, Fig. 9.4).

After further reinforcement of how to make the call for comparative situations, the students were given some practice material. The practice material given to the students had each student use a different sample from the same population as they worked on the same investigative question. However, this had the effect of reinforcing the idea that they could use multiple samples to make the call—an unfortunate side effect that had not been anticipated. Therefore, in a modification to the LT, all the practice material involved the same single sample for all students in the class, reflecting what happens in reality, for each investigative question. The use of multiple samples from the same population was appropriate for developing the understanding of making the call and sampling variability; however, it was not appropriate for subsequent practice as it created an unintended confusion for students.

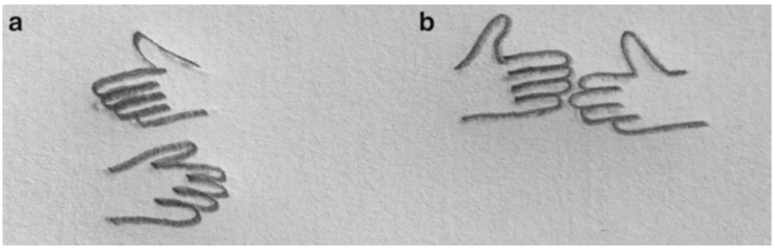


Fig. 9.7 (**a**) Hands close together mimicking two box plots overlapped (on the *left*) and (**b**) hands apart mimicking two box plots with little overlap (on the *right*)

By the end of the LT, based on an analysis of posttest data and individual student interviews, these students were beginning to understand how to make a statistical inference. They were (1) articulating the uncertainty embedded in an inference by drawing upon ideas about sampling variability, (2) making a claim about the population from the sample, and (3) explicitly providing the evidence they used from the data such as distributional shift, overlap, position of the medians, and the decision guide that enabled them to make or not make a call (cf. framework of Makar & Rubin, 2009; Chap. 8). They also seemed to understand how and why the use of the overlap and position of the medians relative to the overlap informed their use of the rule to consistently and coherently make an inference (see Arnold, 2013; Arnold et al., 2011).

9.3.2.4 Discussion and Future Recommendations

Working together to plan the LT and the carefully structured set of learning experiences to support the LT allowed the teacher, Arnold, and wider research team to get a better sense of the possible responses and outcomes for students. Modifications to the LT occurred through extensive debate within the research team, in response to students' difficulties during the lesson, from spontaneous reactions in the classroom to the issue under consideration, through reflection on the lesson or an in-depth analysis of student data after the lessons. The LT for developing the concept of making the call with grade 9 students has been the basis for teacher professional development and subsequent use in their classes.

Defining the learning goal and analyzing the web of concepts are essential ingredients for the construction of LTs. The rich interrelated conceptual repertoire underpinning statistical ideas needs further research including finding ways of developing new conceptual understandings that are not present in current curricula. As this case illustrates, LTs using DBR can assist in the development of new approaches to statistics and in understanding students' reasoning processes. Other topics in statistics need a similar focus to understand teaching and learning processes better, to generate local theories of instruction, and to explore and identify interesting phenomena.

9.3.3 Case Study 3: Preparing Teachers to Develop a Conceptualization of Repeated Sampling for Inference

9.3.3.1 Introduction

The third case presents a LT for assisting adult learners (mostly secondary and post-secondary mathematics and statistics teachers) in conceptualizing repeated sampling approaches to statistical inference, with particular attention to the role of probability models in that conceptualization. The teachers had already been exposed to formal hypothesis techniques. The intent of this case is to illustrate how and why a team of instructors working in real graduate-level classrooms with a designed LT added further learning experiences in response to their observations on the teachers' reasoning processes.

9.3.3.2 The Learning Goals and the Designed Learning Process

The focus of the LT in this case study was to assist teachers in conceptualizing a repeated sampling approach to inference and to consider their learning with this approach. In a repeated sampling approach to inference, students and teachers should be conceiving of the observed outcome (from an observational study or an experimental design) as resulting from a process that is repeatable and that repeating the process may result in a different outcome. Thus, the question becomes: How unusual is what happened in the particular instance that we know about already? In other words, what is the likelihood of a particular outcome occurring if a process is repeated many times?

Lee and Doerr considered learners' use of probability models as essential to conceptualizing a repeated sampling approach to inference. To produce the LT, they considered the research literature and curriculum development in recent years that had focused on understanding inference and using simulation to enact resampling approaches (cf. Fig. 9.1(2 and 3)). For example, Saldanha and Thompson (2002) reported that when students can visualize a simulation process through a three-tier scheme, they develop a deeper understanding of the process and logic of inference. This scheme is centered around "the images of repeatedly sampling from a population, recording a statistic, and tracking the accumulation of statistics as they distribute themselves along a range of possibilities" (p. 261). Lane-Getaz (2006) offered the simulation process model (SPM) to describe the process of using simulation to develop the logic of inference starting with a question in mind, "what if," to investigate a problem including three tiers: population parameters, random samples, and distribution of sample statistics. In line with Lane-Getaz's suggestion, Garfield and Ben-Zvi (2008) and Garfield, delMas, and Zieffler (2012) used a generalized structure to the logic of a simulation approach to inference in their curriculum materials. Their structure includes specifying a model, using the model to generate simulated data for a single trial and then multiple trials, each time collecting a statistic of interest, and finally using the distribution of collected summary measures to compare observed data with the behavior of the model.

Saldanha and Liu (2014) described work with learners in repeated sampling tasks and made the case that students should develop a stochastic conception of an event that "entails thinking of it as an instantiation of an underlying repeatable process, whereas a non-stochastic conception entails thinking of an event as unrepeatable or never to be repeated" (p. 382). Such a stochastic conception includes seeing an event as an expression of some process that could be repeated under similar conditions that produces a collection of outcomes and "reciprocally, seeing a collection as having been generated by a stochastic process" (p. 382). All this research literature fed into the development of the LT (cf. Fig. 9.1(2)) including the influence of the modeling perspective of Lesh and Doerr (2003) and the importance of a careful model development sequence for learners. Such a model development sequence emphasizes how learners develop their own models of a context within a LT.

Lee and Doerr's learning goal was for teachers to develop a stochastic conception of events and a generalizable model that they could use to approach inference situations using a repeated sampling approach and for them to be able to assist others in using such an approach (cf. Fig. 9.1(1a)). This model includes understanding the relationships among the problem situation, physical enactments of sampling, representations of those enactments, computer representations, and the underlying randomization (i.e., the probability models discussed above), the distribution of the statistics of interest, and how to interpret and use such a distribution (a sampling distribution) to make a decision. In order for learners to develop that model (and the entailments needed for teaching that model), they hypothesized that they should be able to make connections to and use the underlying probability model of repeatable actions with unpredictable outcomes.

The initial LT of Lee and Doerr is depicted in Fig. 9.8. This represents the key experiences they felt would lead to a generalizable model for how to use a simulation approach to inference. The key experiences in the trajectory are bolded in the center, while the statistical concepts that should be emphasized at each phase in the trajectory are noted on the right, and pedagogical considerations that could be useful in participants' own teaching practices are noted on the left. Both the statistical ideas that needed to come to the fore and pedagogical issues could help inform the development of teachers' understandings.

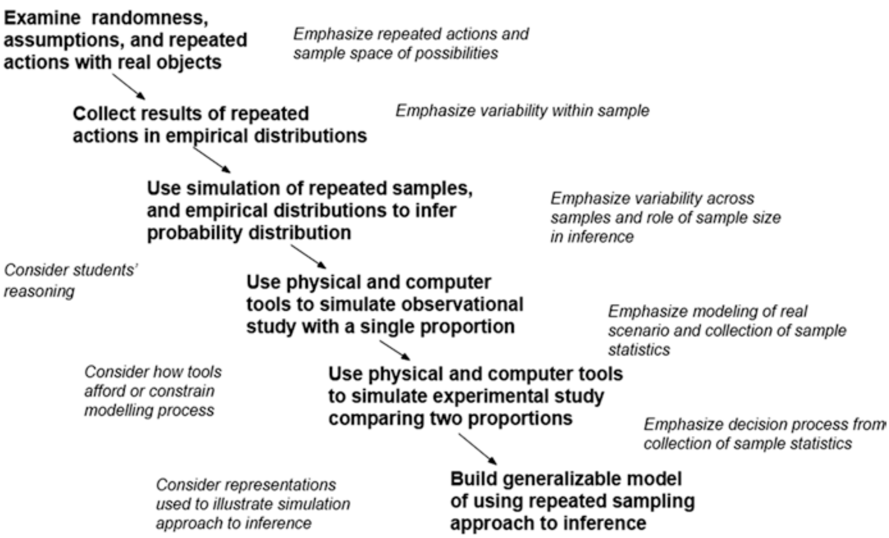


Fig. 9.8 Initial planned LT for a repeated sampling approach to inference

9.3.3.3 The Learning Activities and the Observed Learning Process

Lee and Doerr's research goals were to (1) develop and test a sequence of tasks in a LT that could achieve their learning goals for a particular group of adult learners and (2) identify key conceptualizations that seem to afford a stronger development of a generalized model of repeated sampling approach to inference. The approaches used in DBR (Bakker & van Eerde, 2015), their understanding of the literature on probability models and repeated sampling approaches to inference, and the representations and activities used by others (e.g., Lee, Angotti, & Tarr, 2010) informed their design of the LT (cf. Fig. 9.1(2)). The plan for the initial LT was designed during the 4 months before the course began and then revised during the first 7 weeks of the course as they got to know their learners. The course was taught by 4 instructors (led by Lee and Doerr) over 15 weeks in a once-a-week 3-hour meeting format to 27 teacher participants across 2 institutions.

What follows is a description of the LT at the point where teachers are comparing two proportions, the fifth task, and the consequent adjustments made to the LT based on their ongoing analysis of their learners' successes and struggles.

For the *fifth task*, they wanted teachers to apply their developing repeated sampling model for understanding the likelihood of a single proportion to the comparison of two proportions from an experimental design study (see fourth and fifth bolded goal in the initial LT in Fig. 9.8). They modified the *Dolphin Therapy* task (Catalysts for Change, 2012) to ask teachers to create a by-hand simulation using index *cards* that would answer the question: Can swimming with dolphins be therapeutic for patients suffering from depression? In the experiment, in the dolphin-swimming group (treatment), 10/15 patients improved their depression, while 3/15 improved in the control group. The question is whether that result indicates that swimming with dolphins is therapeutic for depression. The teachers were given 30 index cards marked with results from the study (13 cards marked "YES" for those benefiting with swimming with dolphins, and 17 cards marked "NO").

Lee and Doerr anticipated that how to conceive the random assignment in groups as a repeatable action would not be obvious, an important consideration when designing a LT. A variety of methods were created by teachers. After the discussion to draw out the importance of the assumptions of random assignment and that a patient's outcome does not change regardless of group assigned, the class eventually agreed to shuffle the cards representing the 30 patient outcomes and deal cards into 2 groups of 15. By repeating this action and computing the difference in proportion of YESs, they could examine a distribution of the difference in proportions on a shared class dot plot and consider how likely it is that the benefits of therapy reported in the original study happened by chance alone.

The *Dolphin Therapy* hands-on experience was followed by a *sixth task* that was another model exploration activity where the sampling distribution was explored again in *Statkey* (Lock, Lock, Morgan, Lock, & Lock, 2013) and *TinkerPlots* (Konold & Miller, 2005). Many of their teachers seemed to struggle with the multi-tiered process involved in doing a simulation through repeated sampling for this comparing proportions task. It was sometimes difficult for them to keep in mind all the steps of the process that were happening in the computer. They also struggled with interpreting the sampling distribution in terms of how to use it to make an inference. The *seventh task* provided an opportunity for teachers to further explore the structure of their developing models by reading two articles (Lane-Getaz, 2006; Lee, Starling, & Gonzalez, 2014) in which diagrams were used to illustrate the simulation approach.

In the weekly team meetings, the four instructors (including Lee and Doerr) discussed the teachers' struggles with the repeated sampling approaches used in the two simulation tasks. They were not convinced that their learners had developed a general model for how to use a simulation approach to inference that they could apply to other situations and use for teaching students to use such an approach. Thus, they designed a new *eighth task* to allow teachers an opportunity to express their developing conceptions of the simulation process in terms of how they would help students understand the process. They considered that this task was an opportunity for teachers to explore their representations of the structure of models of repeated sampling for drawing inferences that would serve a pedagogical purpose. That is, the intended audience for this representation would be the future students of the teachers, and this representation hence served a perceived purpose of explaining the structure of models of repeated sampling to other learners. Teachers worked in small groups to do the following:

> Suppose you were going to use a repeated sampling approach with your students to help them use a simulation (with physical objects or computer models) to investigate if an observed statistic is likely or unlikely to occur. Draw a diagram you could use to help students understand the general process used for applying randomization techniques for solving these types of tasks.

Both during class and in the post-class analysis, the instructors noticed the wide variety of representations expressed in teachers' diagrams. Many teachers expressed some aspect of the modeling process from the real-world problem (though not always explicit) and that a collection of statistics is used for examining likelihood; however, their diagrams were much less explicit about the "randomize and repeat" phases in a simulation approach (e.g., see sample diagrams in Fig. 9.9).

Lee and Doerr's analysis of teachers' diagrams and the classroom conversations led to the design of an additional *ninth task* that was structurally similar to the *Dolphin Therapy* task but required an adaptation of their previous model since it involved comparing means for two unequally sized groups. In addition, they deliberately changed the form of the manipulatives (using unmarked flat wooden craft sticks rather than pre-marked index cards) to further push the learners in understanding the role of randomization in their model of repeated sampling. The teachers had varied approaches to recognizing what the repeatable action was in the scenario. Many used the craft sticks in some way, with slight variations from each other, to indicate scores and repeatedly reassigning those scores into two different unequal sized groups. Some teachers really struggled and did not create viable ways of representing the scores or reassignment to groups. Their attempts at applying their

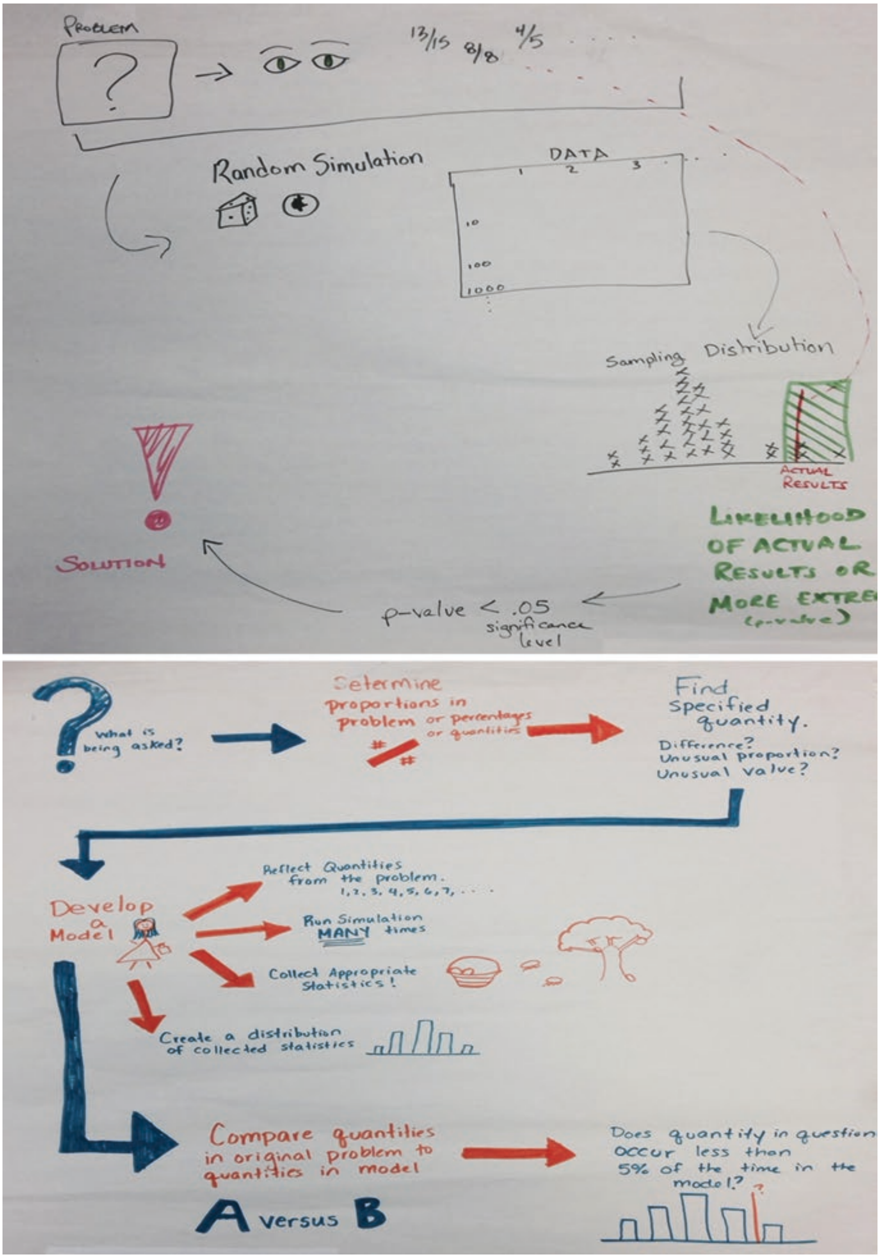


Fig. 9.9 Two samples of teachers' diagrams

model for a repeated sampling approach to inference to create this simulation in such a different context really illuminated the fragility of their models and conceptual understanding.

9.3.3.4 Discussion and Future Recommendations

This case illustrates how an ongoing analysis and instructional experiences impact the development of instructional tasks hypothesized as needed to assist learners in further developing the intended learning goals. Retrospective analysis of learners' work also can be used to modify a LT, in this case for using a simulation approach to inference. This analysis led to a realization that more attention needs to be given to the modeling process, the explicit role of probability in inference, and use of probability language. There is a two-part modeling process that should be made explicit. The first is to create a local specific *model of* the real-world context in statistical terms. The second is creating a simulation process that models the repeatable actions in the original problem and can be used to generate random samples. Most previous works have combined these two aspects into a single "model" or "population" level. There seems to also be a need to be more explicit concerning building a distribution of sample statistics, *viewing the distribution as an empirical probability distribution*, using the distribution to reason about the observed statistic, and making a claim about the chance of that observed statistic occurring. Lee, Doerr, Tran, and Lovett (2016) elaborate on these suggestions. It is important to recall that learners in this case had previous exposure and experience with learning traditional inference techniques, and some had experiences in teaching such techniques. There were only two who had previous experience in using a repeated sampling approach in their own curriculum materials with their students. Thus, the initial LT and sequence of tasks were designed with these learners in mind (cf. Fig. 9.1(3)). Researchers and teachers working with learners first engaging with inference through repeated sampling will need to adapt and adjust the LT as needed.

The LT discussed in this case study demonstrates how LTs are useful for identifying and exploring learners' reasoning processes, building new conceptual approaches for learning statistics, contributing to the research knowledge base, and directing the focus of future research.

9.4 Conclusion

LTs have been critical in the development of statistics education research and in enhancing students' learning in the classroom. LTs are not just a sequence of lessons; rather they are deliberately planned and modified based on careful analyses of the research literature, the web of concepts underpinning the learning goal, and the student responses. This chapter has focused on researchers using LTs, but we recommend that teachers, as action researchers in their own classroom, use LTs to understand and improve their students' learning. Additionally, we recommend that teachers co-design LTs with other teachers to reflect the intentions of their curriculum and the realities of their classrooms (see Chap. 16). Co-designing LTs with researchers is also a possibility. We now reflect on what we can learn from the case studies and then propose four recommendations.

9.4.1 Reflection on the Case Studies

The LTs in the three case studies shared many commonalities. At a meta-level, all shared LTs that combined, in an interactive process, curriculum development and research and sequences of tasks and supporting students' thinking and performance. Furthermore, there was collaboration with teachers in classroom settings reflecting a participationist research paradigm (Sfard, 2005). Differences existed depending on the purpose of the research, the existing research literature, and how many cycles of teaching experiments were implemented. All the case studies, however, reflected the LT iterative process outlined in Fig. 9.1 and the components necessary to inform its design.

All the studies started with a problem. Case Study 1 sought to model students' understanding of variability over time. Case Study 2 had a defined goal of making the call when comparing two box plots and then ascertained the myriad of concepts that underpinned making a judgment under uncertainty. Case Study 3 began with the researchers' knowledge of the literature on statistics, probability, and modeling and their belief that teachers needed to conceptualize the links among them into a general model. In line with the other two studies, Case Study 3 developed a hypothetical learning process that aimed to scaffold teachers' thinking toward a general model realization about repeated sampling for making an inference. Case Studies 1 and 3 drew on some existing learning activities for their LTs, whereas Case Study 2 invented its own. Whether inventing new tasks for LTs or not, all attended to delineating the statistical big ideas and concepts underpinning the learning goal and strived to engage students in the LT's defined abstract notions using innovative learning approaches. During teaching, as students engaged with the learning tasks, their actions, representations, and thinking were observed and analyzed. Consequently there was a feeding forward and back into the LTs, which were modified and altered from the planned LT. Case Study 3 illustrated the importance of a retrospective analysis whereby the researchers, in response to the teachers' fragile understanding of models for repeated sampling in inference, proposed some new key conceptualizations.

Compared to research that gauges levels and types of thinking based on survey questions or explicating students' thought processes when engaging with several tasks, research that uses LTs and DBR methodology has the potential to have more impact on learning in classrooms as Case Studies 1 and 2 show (see Chap. 16 also). While acknowledging that the findings from the former type of research are vital for the designing of LTs, the latter type of research is also good at identifying gaps in students' thinking and new avenues to explore (e.g., Case Study 3). A LT can be just one lesson or cover many lessons, but as these studies illustrated, statistical big ideas and concepts take time to experience and take root in students' cognitive infrastructure.

In a critique of LTs used in research, Baroody, Cibulskis, Lai, and Li (2004) believed some of them were overly prescriptive and detailed and consequently an inquiry-based investigative approach was lost. They conjectured that LTs "could be more comprehensible and useful to practitioners if they focused on how big ideas evolve" (p. 253). These case studies did focus on the big ideas and how these might evolve at particular levels, but there is a danger that microanalyses of students' thinking, while important to research, may lead to a plethora of types and levels of reasoning resulting in researchers and teachers using step-by-step procedures in LTs to achieve the learning goals. When designing LTs, an important criterion to consider is the degree of openness permitted in the learning process so as not to lose the investigative spirit inherent in the statistical enterprise and the process of inquiry that is central to statistical thinking and learning (cf. Chap. 10).

The statistical inquiry investigative cycle is the centerpiece of some new curricula (e.g., Ministry of Education, 2007) with students learning how to be "data detectives." As part of enculturating (Garfield & Ben-Zvi, 2008) students into statistical thinking and inquiry (see Chaps. 4 and 7), the development of concepts is essential as well as the development of coherent conceptual infrastructure across the curricula levels. These LTs illustrated how conceptual understanding might be built up in students and teachers. However, researchers may need to remind themselves not to lose sight of the big ideas and the inquiry-based investigative approach when designing LTs. That is, there is a balance between concept-focused and inquiry-based LTs.

9.4.2 Recommendations and Implications

We have four recommendations for future research regarding LTs:

1. Continue exploratory research on LTs of specific topics in statistics.
2. Scale LTs to many diverse classrooms.
3. Build coherent conceptual pathways across curricula and grade levels.
4. Attend to analysis of web of concepts, task design, and methods of data analysis.

Much of the research using LTs has been within one topic domain at one curriculum level with a few groups of students. As Case Studies 2 and 3 showed, exploratory research with one group of students that either treads into new territory or investigates a concept from a new angle can provide invaluable insights into garnering understanding about teaching and learning processes. These small-scale studies can facilitate the generation of more refined local theories about teaching and learning certain topics in statistics. Thus, our *first recommendation* is that researchers continue using LTs in their research as they have enormous potential to explore and identify interesting phenomena and to develop theories about learning.

The *second recommendation*, which Case Study 1 attempted to address, is scalability to many classrooms. The challenge for Case Study 1 was accurately capturing typical responses, describing them in terms of increasing levels of sophistication, and communicating these ideas effectively to teachers. Another open question was how to make teachers aware of LT research results so that they could anticipate the possible student ideas and challenges, provide opportunities for ideas to emerge, and then use data on student learning to support continued progress in learning. Such challenges and questions will need to be addressed when expanding successful LTs to a broad range of classrooms. When LTs are considered to have the potential to be shared, we recommend that researchers think about collaborating in new research projects to address how to manage implementation on a larger scale. Where necessary, researchers may need to alter their LTs in response to new findings as a result of more people such as curriculum developers, professional development facilitators, and teachers being involved in the implementation (e.g., Lehrer et al., 2014).

The *third recommendation* is building curriculum coherence for teachers and students across the grade levels. What is needed is a major collaboration of researchers worldwide to work out the big ideas and web of concepts that have been researched and where more research is needed (e.g., covariation). They could then attempt to map across the curriculum the main conceptual pathways and identify the LTs that exist and may be used given the time constraints of curricula. We recommend, as a start, that researchers using LTs could devise and research a pathway for growing students' knowledge and thinking in one topic domain from grades 1 to 12 in a similar vein to Case Study 1 with its learning maps, relational learning clusters, and big ideas for grade 6. Perceiving across the curriculum, an evolving conceptual pathway together with LTs toward a big idea could be useful for curriculum developers and for the research community.

In Sect. 9.2 we identified three aspects regarding the design and use of LTs that seemed to need more attention in research. Hence, our *fourth recommendation* is that researchers conduct more in-depth analyses of the web of concepts underpinning their learning goal, carefully consider the literature on task design and the influence the task will have on students' learning, and devise more transparent ways of analyzing data gathered and providing evidence, particularly for classroom interactions. Also meta-LT research is required to study LTs as a methodological tool. Addressing these issues, which seem to be currently missing in statistics education research using LTs, would move the field forward.

In statistics education research, the use of LTs as an instrument in DBR has resulted in a fecund route for learning about students' thinking and has opened up many new challenges and avenues for future research. As technology changes approaches to learning, there is now an even greater need to focus on the big ideas and concepts that will endure despite those changes. We believe that using LTs and DBR will continue to provide a fruitful and rewarding pathway for future researchers.

References

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal, 32*(1), 23–38.

Arnold, P. (2013). Statistical investigative questions – An enquiry into posing and answering investigative questions from existing data. (Doctoral thesis). Retrieved from https://researchspace.auckland.ac.nz/handle/2292/21305

Arnold, P., Pfannkuch, M., Wild, C. J., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands-on to "movies". *Journal of Statistics Education, 19*(2), 1–32.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: Freudenthal Institute.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer.

Bakker, A., & van Eerde, D. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 429–466). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/978-94-017-9181-6_16

Baroody, A., Cibulskis, M., Lai, M., & Li, X. (2004). Comments on the use of learning trajectories in curriculum development and research. *Mathematical Thinking and Learning, 6*(2), 227–260. https://doi.org/10.1207/s153277833mtl0602_8

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal, 3*(2), 42–63.

Ben-Zvi, D., & Garfield, J. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 397–409). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction, 14*(3), 325–341. https://doi.org/10.1016/j.learninstruc.2004.06.006

Catalysts for Change. (2012). *Statistical thinking: A simulation approach to modeling uncertainty*. Minnesota, MN: Catalyst Press.

CensusAtSchool New Zealand. (2003). Retrieved from www.censusatschool.org.nz

Clements, D., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning, 6*(2), 81–89.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics. (College- and Career-Readiness Standards and K12 standards in English Language Arts and Math*). Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. Retrieved from http://www.corestandards.org

Confrey, J. (1991). Learning to listen: A student's understanding of powers of ten. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education* (pp. 111–138). Dordrecht, The Netherlands: Kluwer Academic Press.

Confrey, J. (2002). Sixth grade pre-algebra curriculum at St. Francis School, Austin, TX. Unpublished document.

Confrey, J. (2006). The evolution of design studies as methodology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 135–151). Cambridge, UK: Cambridge University Press.

Confrey, J. (2015). Some possible implications of data-intensive research in education—The value of learning maps and evidence-centered design of assessment to educational data mining. In C. Dede (Ed.), *Data-intensive research in education: Current work and next steps. Report on two National Science Foundation-sponsored Computing Research Association workshops* (pp. 79–87). Washington, DC: Computing Research Association.

Confrey, J., & Lachance, A. (2000). Transformative teaching experiments through conjecture-driven research design. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 231–265). Mahwah, NJ: Lawrence Erlbaum Associates.

Design-Based Implementation Research. (2016). Retrieved from http://learndbir.org/

Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*(1&2), 132–151. https://doi.org/10.1080/10986065.2011.538294

Finzer, B., Konold, C. & Erickson, T. (2012). Data games. Retrieved from http://play.ccssgames.com

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report*. Alexandria, VA: American Statistical Association.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal, 4*(1), 92–99.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372–396.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, The Netherlands: Springer.

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM Mathematics Education, 44*(7), 883–898.

Gravemeijer, K. (2004). Local instruction theories as means of support for teachers in reform mathematics education. *Mathematical Thinking and Learning, 6*(2), 105–128.

Gravemeijer, K., & Cobb, P. (2006). Design research from a learning design perspective. In J. van den Akker (Ed.), *Educational design research* (pp. 17–51). London: Routledge.

Konold, C., & Miller, C. (2005). *TinkerPlots dynamic data exploration: Statistics software for middle school curricula*. Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Lane-Getaz, S. J. (2006). What is statistical thinking and how is it developed? In G. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth NCTM yearbook* (pp. 272–289). Reston, VA: National Council of Teachers of Mathematics.

Lee, H. S., Angotti, R. L., & Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal, 9*(1), 68–96.

Lee, H. S., Doerr, H. M., Tran, D., & Lovett, J. N. (2016). The role of probability in developing learners' models of simulation approaches to inference. *Statistics Education Research Journal, 15*(2), 216–238.

Lee, H. S., Starling, T. T., & Gonzalez, M. D. (2014). Connecting research to practice: Using data to motivate empirical sampling distributions. *Mathematics Teacher, 107*(6), 465–469.

Lehrer, R. (2016). Data modeling. Retrieved from http://modelingdata.org/

Lehrer, R., & Kim, M. J. (2009). Structuring variability by negotiating its measure. *Mathematics Education Research Journal, 21*(2), 116–133.

Lehrer, R., Kim, M. J., Ayers, E., & Wilson, M. (2014). Toward establishing a learning progression to support the development of statistical reasoning. In A. Maloney, J. Confrey, & K. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 32–59). Charlotte, NC: Information Age Publishers.

Lehrer, R., Kim, M. J., & Jones, R. S. (2011). Developing conceptions of statistics by designing measures of distribution. *ZDM Mathematics Education, 43*(5), 723–736.

Lehrer, R., & Schauble, L. (Eds.) (2002). *Investigating real data in the classroom: Expanding children's understanding of math and science*. New York, NY: Teachers College Press.

Lesh, R. A., & Doerr, H. M. (Eds.). (2003). *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lock, R., Lock, P., Morgan, K., Lock, E., & Lock, D. (2013). StatKey. Retrieved from http://lock5stat.com/statkey

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 152–173.

Makar, K., & Confrey, J. (2005). Variation talk: Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27–54.

Makar, K., & Rubin, A. (2009). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 8*(1), 82–105.

Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, New Zealand: Learning Media.

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*(2&3), 131–156.

Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education, 2*(3), 149–167.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17–46). Dordrecht, The Netherlands: Kluwer.

Prediger, S., Gravemeijer, K., & Confrey, J. (2015). Design research with a focus on learning processes – An overview on achievements and challenges. *ZDM Mathematics Education, 47*(6), 877–891.

Radford, L. (2009). Why do gestures matter? Sensuous cognition and the palpability for mathematical meanings. *Educational Studies in Mathematics, 70*(2), 111–126.

Rubin, A., Bruce, B., & Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the 3rd International Conference on Teaching Statistics, Dunedin, New Zealand* (Vol. 2, pp. 314–319). Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/18/BOOK1/A9-4.pdf

Saldanha, L., & Liu, Y. (2014). Challenges of developing coherent probabilistic reasoning: Rethinking randomness and probability from a stochastic perspective. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 367–396). Dordrecht, The Netherlands: Springer.

Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257–270.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184.

Sfard, A. (2005). What could be more practical than good research? On mutual relations between research and practice of mathematics education. *Educational Studies in Mathematics Education, 58*(3), 393–413.

Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114–145.

Watson, A., & Ohtani, M. (Eds.). (2015). *Task design in mathematics education*. New York, NY: Springer.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–248.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. (2011). Towards more accessible conceptions of statistical inference (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2), 247–295.

Chapter 10
Research on Statistics Teachers' Cognitive and Affective Characteristics

Randall Groth and Maria Meletiou-Mavrotheris

Abstract Research about statistics teachers faces a unique challenge. It is not sufficient to account only for teachers' cognition and affect in regard to the subject matter of statistics. We also need to understand the personal characteristics teachers have related to developing the statistics-related cognitive and affective traits of students. Toward this end, researchers have supplemented studies of teachers' subject matter knowledge with studies of their pedagogical content knowledge, technological pedagogical statistical knowledge (TPSK), beliefs, and attitudes relevant to teaching statistics. We describe existing models and empirical research concerning each of these characteristics. Written assessments, interview techniques, and observation methods for assessing teachers' development of the characteristics are described as well. Strengths and limitations of existing models and assessments are discussed. We conclude by summarizing statistics teacher education research in the specific areas of data, uncertainty, and statistical inference. We close with recommendations about how statistics teachers' cognitive and affective characteristics may be developed by learning from teaching practice, immersion in statistical content, and use of technological environments. Opportunities and directions for future research appear throughout the chapter. Some specific research needs include progressive development of improved models for statistics teachers' cognition and affect along with robust qualitative and quantitative assessment tools.

Keywords Affect • Assessment techniques • Cognition • Pedagogical content knowledge • Subject matter knowledge • Teaching statistics • Technological pedagogical statistical knowledge (TPSK)

R. Groth (✉)
Education Specialty Department, Salisbury University, Salisbury, MD, USA
e-mail: regroth@salisbury.edu

M. Meletiou-Mavrotheris
School of Humanities, Social and Education Sciences, European University Cyprus, Egkomi, Cyprus
e-mail: m.mavrotheris@euc.ac.cy

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_10

10.1 Introduction

The previous chapters in this handbook have dealt with the nature of statistics and students' knowledge of the discipline. We now turn our attention to important mediators between students and the discipline: statistics teachers (primary, secondary, and tertiary). Teachers can be considered the third vertex in a didactic triangle, with students and content as the other two vertices (Goodchild & Sriraman, 2012). The mediating role of the teacher has motivated researchers to study teachers' cognitive and affective characteristics. Certainly, not all teacher and student interactions can be fully explained by these characteristics alone, since contextual constraints such as supervisor expectations, institutional policies, and instructional resources available are influential (Forgasz & Leder, 2008; Sullivan & Mousley, 2001). Diversity among students and equity concerns also come into play (Cobb, 1999). Nonetheless, research suggests that understanding teachers' individual characteristics is an essential part of studying teachers' impact on students' learning. For example, student achievement in statistics has been found to be positively associated with teachers' knowledge (Callingham, Carmichael, & Watson, 2016).

In this chapter, we focus specifically on research findings related to the cognitive and affective characteristics of statistics teachers. Precise definitions of "cognition" and "affect" are elusive in the literature, but we conceive of cognitive characteristics as being related to the knowledge and reasoning processes needed for teaching statistics and affective characteristics as being related to dispositions, emotions, attitudes, and beliefs about teaching statistics (McLeod, 1992). In many cases, it is difficult to separate cognition from affect. Beliefs, for example, though often discussed in connection with affect, are widely acknowledged to have cognitive components as well (Philipp, 2007). Hence, in this chapter, our primary goal is not to separate characteristics neatly into "cognitive" and "affective" bins, but rather to identify characteristics that may help shed light on the nature of teachers' mediating role between students and statistics.

We begin the chapter with descriptions of salient theoretical constructs related to statistics teachers' cognition (Sect. 10.2) and affect (Sect. 10.3). We then turn to methods for assessing attainment of these constructs (Sect. 10.4). Finally, we summarize findings from research in regard to the constructs (Sect. 10.5). In doing so, we seek to portray the current state of the art and identify fruitful directions for further research (Sects. 10.6 and 10.7).

10.2 Constructs for Describing Teachers' Cognitive Characteristics

Researchers employ various theoretical models to study cognition related to teaching statistics. These models generally acknowledge that knowing statistics is a necessary, but not sufficient, condition for teaching it. This resonates with Shulman's

(1987) influential assertion that teachers need pedagogical content knowledge, which is a "special amalgam of content and pedagogy that is uniquely the province of teachers" (p. 8). Building on Shulman's work, the Learning Mathematics for Teaching (LMT) Project conceptualized content knowledge for teaching as consisting of both subject matter knowledge and pedagogical content knowledge (Ball, Thames, & Phelps, 2008; Hill, Ball, & Schilling, 2008). Several studies of statistical knowledge for teaching (SKT) have profitably used adaptations of the LMT model (e.g., Burgess, 2011; González, 2014; Groth, 2013; Leavy, 2015; Noll, 2011; Wassong & Biehler, 2010). Hence, we describe possible elements of subject matter knowledge (Sect. 10.2.1) and pedagogical content knowledge (Sect. 10.2.2) related to the LMT model next.

Although the LMT model appears frequently in statistics teacher education research, it would be inaccurate to portray it as the only model employed. We will also describe work that challenges the field to continue to think critically about the precise nature of the elements of SKT, their relationships with one another, and how they develop (Sect. 10.2.3). We close with a discussion of technological pedagogical content knowledge and how it relates to other research on teachers' cognitive characteristics (Sect. 10.2.4).

10.2.1 Subject Matter Knowledge

Subject matter knowledge can be conceptualized as having three sub-domains: common content knowledge, specialized content knowledge, and horizon knowledge (Ball et al., 2008; Hill et al., 2008).

Hill et al. (2008) described common content knowledge as "knowledge that is used in the work of teaching in ways in common with how it is used in many other professions or occupations that also use mathematics" (p. 377). At the university level, prospective teachers often study aspects of common statistical knowledge alongside those preparing for other professions. For example, knowing how to compute and interpret descriptive statistics such as mean, median, and interquartile range is valuable both to teachers and to other professionals (Groth, 2007).

Specialized content knowledge can be described as "the mathematical knowledge that allows teachers to engage in particular *teaching* tasks, including how to accurately represent mathematical ideas, provide mathematical explanations for common rules and procedures, and examine and understand unusual solution methods to problems" (Hill et al., 2008, p. 378). Specialized knowledge of statistics might involve knowing how to represent the mean as a typical value, a fair share, and a signal amid noise (Wassong & Biehler, 2010). It might also entail the ability to analyze students' statistically naïve interpretations of data (Burgess, 2011). Similarly, appraising novel student-invented graphical representations may be done by drawing upon specialized knowledge (Groth, 2013).

Ball and Bass (2009) spoke of horizon knowledge as that which allows teachers to see connections between content studied at a particular grade level and major disciplinary structures, ideas, practices, and sensibilities. Consider the case of standard deviation. Seventh-grade US teachers using the Common Core State Standards (National Governors Association for Best Practices & Council of Chief State School Officers, 2010) would not teach this idea directly to their students, but they would be responsible for teaching the related idea of mean absolute deviation (MAD). Teachers might conceive of the MAD in at least two different ways: (1) as an algorithm for students to compute and master or (2) as a precursor for the study of standard deviation (Groth, 2014). Teachers holding the latter view would seem more likely to select tasks and ask questions that lead toward the concept of standard deviation that is on the statistical horizon. Teachers holding the former view may not be able to imbue their instruction with this perspective, and they may reduce the study of the MAD to rote learning of a procedure. Many questions about horizon knowledge remain for exploration by researchers, such as: (1) What specific aspects does horizon knowledge entail? (2) How is horizon knowledge best developed? (3) What can teachers with well-developed horizon knowledge do for students that others cannot?

Research involving teachers frequently focuses on their subject matter knowledge. Specific findings regarding the nature of different elements of teachers' subject matter knowledge appear in Sect. 10.5.

10.2.2 Pedagogical Content Knowledge

Hill et al. (2008) hypothesized that pedagogical content knowledge has three subdomains: knowledge of content and students, knowledge of content and teaching, and curriculum knowledge.

The first sub-domain, knowledge of content and students, pertains to teachers' knowledge of students' thinking patterns and problem-solving strategies (Hill et al., 2008). The importance of this type of knowledge is well established in the literature. Teachers participating in professional development about how students think about content tend to be more effective in facilitating students' learning (Franke, Kazemi, & Battey, 2007). Statistics education researchers have hypothesized that knowledge of content and students for statistics may consist of elements such as understanding students' difficulty learning the mean conceptually (Wassong & Biehler, 2010), comprehending student difficulties sorting data (Burgess, 2011), and knowing differences between how students tend to read dot plots and box plots (Groth, 2013). Comprehensively conceptualizing the nature of knowledge of content and students for statistics and its impact on student learning constitute important tasks for future research.

Knowledge of content and teaching is a combination of knowing about teaching and knowing about subject matter. It can help teachers with tasks such as choosing models and examples that bring out important aspects of content (Ball et al., 2008). It appears that knowledge of content and students contributes to knowledge of content and teaching. Consider, for example, a teacher who knows that transitioning from dot plot displays to box plots and histograms is a difficult task for students. Such a teacher is in position to select tasks to help students gradually transition from one display to the next (Groth, 2013).

Ball and Bass (2009) spoke of curriculum knowledge as including knowledge of educational goals, standards, and grade levels where specific concepts appear. This type of knowledge may help teachers appropriately sequence the introduction of statistical ideas in a curriculum (Godino, Ortiz, Roa, & Wilhelmi, 2011). However, there is considerable variability in how teachers interpret curriculum materials. When given a curriculum, some implement it with a high amount of fidelity to the curriculum authors' intentions, and others do not (Tarr et al., 2008). Sometimes this degrades the quality of instruction, but other times may help improve it (Brown, Pitvorec, Ditto, & Kelso, 2009). Hence, carefully examining statistics teachers' curriculum knowledge has the potential to help explain underlying reasons for instructional dynamics observed in the classroom.

Although pedagogical content knowledge appears as a separate category from content knowledge in the LMT framework, in practice, it is difficult, and often not advisable, to separate the two. Hence, in the summary of research appearing in Sect. 10.5, we report findings about teachers' pedagogical content knowledge, for the most part, alongside subject matter knowledge findings.

10.2.3 Continuing the Work of Precisely Defining SKT Elements, Their Relationships to One Another, and Their Development

Given the preceding description of the elements of SKT, some may gain the impression that it is a static trait rather than one that evolves and changes continuously within classroom contexts. LMT-based models are sometimes perceived in this manner (Venkat & Adler, 2014). Some theoretical work serves to cast SKT in a more dynamic light.

Working from the LMT framework and empirical data, Groth (2013) theorized about processes involved in individuals' transformation of statistics subject matter knowledge into forms that are useful for teaching. Central to the analysis is the idea that teachers' key developmental understandings of subject matter knowledge (Simon, 2006) are, alone, not sufficient for teaching. Teachers who have key developmental understandings must also learn to view subject matter knowledge from students' perspectives in order to create pedagogically powerful ideas (Silverman & Thompson, 2008). In terms of the LMT framework, this suggests that knowledge of content and students is a precursor to developing knowledge of content and teaching. That is, teachers should understand students' learning needs in order to design and select teaching methods suitable for addressing them. The potential link between knowledge of content and students and knowledge of content and teaching deserves more research attention, as it is difficult to conceive of a teacher with robust knowledge of content and teaching but underdeveloped knowledge of content and students. Investigating whether one sort of knowledge is usually prerequisite to another could help effectively sequence learning experiences for teachers.

Studies in the traditions of design research (Bakker & van Eerde, 2015) and didactic engineering (Artigue, 2015) may help further reveal dynamic processes involved in SKT development. Such studies involve iterative cycles of research and development of instructional sequences in classroom settings. As the cycles occur, data yield contextually rich information about teachers' knowledge and its enactment in practice. Working from the perspective of teachers' knowledge situated in an institutional context, Godino et al. (2011) proposed facets of professional knowledge for teaching statistics that differ from those in the LMT model. One such facet was that teachers need knowledge of "students' attitudes, emotions, and motivations regarding the content" (p. 279). This type of knowledge is similar to knowledge of content and students in its focus on student characteristics, but different in that it deals with the importance of knowing children's affect in regard to statistics (and not just children's statistical cognition). It would be profitable for researchers to take advantage of different conceptualizations of the nature of SKT as starting points for comparing and contrasting viewpoints. Done systematically, such theoretical comparisons could lead to the incremental development of progressively more sophisticated models of SKT and how it develops.

In any model of SKT that is constructed, it is important for researchers to acknowledge that mathematics and statistics are distinct disciplines. Mathematics and statistics differ in their "origins, subject matter, foundational questions, and standards" (Moore, 1988, p. 3). Therefore, it is reasonable to assume that the knowledge needed for teaching statistics is not precisely equivalent to the knowledge needed for teaching mathematics (Groth, 2007). Hence, as theoretical work on the conceptualization of SKT continues, researchers must be careful to distinguish, as necessary, between professional knowledge needed for teaching mathematics and that needed for teaching statistics.

10.2.4 Technological Pedagogical Content Knowledge

Shulman's (1987) notion of pedagogical content knowledge is the basis for another related, yet somewhat distinct, body of research on teachers' knowledge. As digital technologies became more prevalent in classrooms, it was apparent that teachers needed *technological pedagogical content knowledge* (TPCK) to effectively use them for instruction (Koehler & Mishra, 2008). TPCK is a complex interaction among knowledge of content, pedagogy, and technology. Some theoretical work to conceptualize TPCK for statistics appears in this section, and some work to help develop teachers' TPCK appears in Sect. 10.6.3.

Lee and Hollebrands (2011) offered a framework to operationalize TPCK for statistics. Their framework posits statistical knowledge as the basis for technological statistical knowledge (TSK). TSK is a blend of technology knowledge and statistics knowledge. TSK includes knowledge of technologies that are both *amplifiers* and *reorganizers* (Ben-Zvi, 2000; Lee & Hollebrands, 2008). Amplifiers help automate processes that could be done by hand, such as computing least-squares regression lines (Lee & Hollebrands, 2008). Reorganizers "extend what teachers may be able to do without technology to help students reorganize and change their statistical conceptions" (Lee & Hollebrands, 2008, p. 329). For instance, TinkerPlots (Konold & Miller, 2011) affords the opportunity to generate and link multiple graphical representations. Using TinkerPlots to produce suitable representations for data is another activity that engages TSK (Lee et al., 2014).

TSK must ultimately merge with pedagogical knowledge if teachers are to develop technological pedagogical statistical knowledge (TPSK). An example of a task requiring TPSK was discussed by Wilson, Lee, and Hollebrands (2011). Teachers used pedagogical, statistical, and technological knowledge in analyzing a video case of children working with TinkerPlots. To analyze the case, teachers attended to how students thought about statistical tasks, how they used TinkerPlots in solving them, how the technology assisted or hindered students' learning, and the strengths and weaknesses of the task given to students. Statistical knowledge, or even TSK, would not be sufficient for analyzing these elements of the case.

At present, the literature provides a more comprehensive portrait of teachers' TSK than it does TPSK, though investigation of both types of knowledge is in its beginning stages. Also requiring research attention are the potential links among statistical knowledge, TSK, and TPSK. The Lee and Hollebrands (2011) framework provides a starting point for such investigations, but Lee et al. (2014) acknowledge that empirical work remains to be done to test the conjecture that teachers' TSK impacts their TPSK and teaching practices.

10.3 Constructs for Describing Teachers' Affective Characteristics

In the affective domain, beliefs and attitudes of statistics teachers have received research attention. However, the terms "beliefs" and "attitudes" are not used uniformly across studies. Philipp (2007) encountered the same dilemma in writing about beliefs and attitudes related to mathematics. To address the problem, he offered general descriptions that capture much of what authors often mean when using the two terms:

- Attitudes: "manners of acting, feeling, or thinking that show one's disposition or opinion. Attitudes change more slowly than emotions, but they change more quickly than beliefs" (p. 259).

- Beliefs: "Psychologically held understandings, premises, or propositions about the world that are thought to be true. Beliefs are more cognitive, are felt less intensely, and are harder to change than attitudes" (p. 259).

These characterizations provide starting points for our discussion of statistics teachers' beliefs (Sect. 10.3.1) and attitudes (Sect. 10.3.2).

10.3.1 Beliefs

Research describes several types of beliefs connected to teaching statistics. These include beliefs about the relationship between mathematics and statistics, goals and strategies for statistics instruction, and self-efficacy to teach statistics.

10.3.1.1 Beliefs About the Relationship Between Mathematics and Statistics

Statistics is often taught as part of a mathematics curriculum or in a mathematics department. This arrangement can support the belief that statistics is a branch of mathematics rather than a discipline in its own right (Burrill & Biehler, 2011). Rossman, Chance, and Medina (2006) argued that this is not a useful belief for teachers to hold, as it may lead to lack of instructional emphasis on the nature and role of context, measurement, data collection, and uncertainty in statistics. Similar concerns have been expressed by others (Gattuso, 2008; Scheaffer, 2006). Empirical data lend support to the validity of such concerns. Begg and Edwards (1999) found that teachers tended to acknowledge the cross-curricular nature of statistics yet still generally taught it as a unit of mathematics. Yang (2014) suggested that teacher beliefs about the differences between statistics and mathematics may be influenced by national curricula and assessments and that it would be worthwhile to explore the influence of these factors.

10.3.1.2 Beliefs About Goals and Strategies for Statistics Instruction

Eichler (2007) provided an empirically grounded framework for characterizing teachers' beliefs about the goals of statistics instruction. The framework included four categories of beliefs: traditionalist, application preparer, everyday life preparer, and structuralist. Traditionalists focus on the study of probability and algorithms in the abstract and not on applications. Application preparers value teaching students the interplay between theory and applications, focusing on the use of algorithms to solve real-world problems. Everyday life preparers take the focus on applications a step further, believing that the study of statistics should be driven by applications rather than theory. Structuralists focus heavily on probability theory, mathematical structure, and algorithms. Structuralists differ from traditionalists in that they believe in using applications as the basis for instruction. Structuralists differ from other groups in that their primary goal is to help students abstract mathematical structure from the applications rather than apply mathematical principles to make sense of situations students encounter outside of school.

Aspects of Eichler's framework resonate with other researchers' findings in regard to teachers' beliefs about strategies and goals of statistics instruction. Sedlmeier and Wassner (2008) found that teachers believed it to be valuable to relate statistics content to daily issues (similar to everyday life preparers), but did not believe in placing as much emphasis on student data gathering or student interests. Pierce and Chick (2011) found that some teachers believe in teaching procedures first and then using applications merely to try to make the procedures more interesting. Such a strategy may reflect application preparer and/or structuralist tendencies. Comprehensively mapping the relationships between observed teaching strategies and specific beliefs about the goals of statistics instruction is an interesting empirical task for which some infrastructure currently exists, and it awaits additional research attention.

10.3.1.3 Self-Efficacy Beliefs About Teaching Statistics

Harrell-Williams, Sorto, Pierce, Lesser, and Murphy (2014) argued that it is important to measure self-efficacy to teach statistics. Teacher self-efficacy can be defined as a teacher's belief that he or she has the ability to bring about student learning (Ashton, 1985). Harrell-Williams et al. synthesized existing research to conclude that self-efficacy influences teachers' choices of instructional techniques and students' learning. They argued that it is particularly important to consider self-efficacy in regard to teaching statistical investigations. Such a domain-specific portrait of teacher self-efficacy is potentially more informative to teacher educators than more generic assessments.

10.3.2 Attitudes

There is a voluminous body of research on individuals' attitudes toward statistics (Nolan, Beran, & Hecker, 2012), but literature about *teachers'* attitudes toward statistics is more sparse (Estrada, Batanero, & Lancaster, 2011). Available research suggests that teachers tend to value statistics as a subject but find it difficult to enjoy, teach, and learn (Estrada, Batanero, Fortuny, & Díaz, 2005; Martins, Nascimento, & Estrada, 2012). Teachers' attitudes toward statistics are potentially important because they are hypothesized to relate to their persistence in gaining statistical knowledge (Estrada et al., 2005) and willingness to teach the subject (Leavy, Hannigan, & Fitzmaurice, 2013). Teachers' attitudes toward statistics are hypothesized to influence their knowledge of statistics, their teaching practices, and their students' attitudes

(Martins et al., 2012). Several attitude-related hypotheses, however, await strong empirical support. In studies of the impact of attitudes on teachers' content knowledge, for example, researchers have found moderate to low correlations (Hannigan, Gill, & Leavy, 2013; Nasser, 2004). Negative attitudes toward statistics appear to be clearly detrimental (Onwuegbuzie, 2000), but there seems be a limit on the extent to which positive attitudes relate to increased knowledge (Hannigan et al., 2013).

It appears that the field has not yet made a strong distinction between "teachers' attitudes toward statistics" and "attitudes toward teaching statistics." Although assessment items about attitudes toward teaching statistics have been included in some research studies (Martins et al., 2012; Pierce & Chick, 2011), many studies of teachers' attitudes have used instruments intended to measure the attitudes of the general population (Estrada et al., 2011; Hannigan et al., 2013). This might explain why empirical evidence about the impact of attitudes is elusive. If, for example, the field were to systematically conceptualize and investigate teachers' attitudes toward *pedagogical elements* such as statistics curriculum, children's statistical learning, and technology for teaching statistics, might we better understand the impact of teachers' attitudes on statistics teaching and learning?

10.4 Methods for Assessing Statistics Teachers' Cognition and Affect

Assessments of cognition and affect related to teaching statistics come in a variety of forms, spanning the spectrum of written assessments, interviews, and observations. Many studies make use of more than one type of assessment and may involve more than one aspect of teachers' cognition and affect. Below, a representative sample of assessments is discussed. Due to the scope of the chapter, we focus on assessments specifically designed for teachers rather than general standardized scales of cognition and affect that are sometimes used as part of research with teachers.

10.4.1 Written Assessments

Written assessments are often the most practical way to gather information from large groups of teachers. One such assessment, the Diagnostic Teacher Assessment of Mathematics and Science, includes a separate scale of multiple-choice and open-ended items for statistics (Saderholm, Ronau, Brown, & Collins, 2010). The LMT project also designed a scale of multiple-choice items specific to teaching statistics (G. Phelps, personal communication, June 11, 2010). An international comparison of teacher education, the Teacher Education and Development Study in Mathematics, included some items on pedagogical content knowledge for statistics among items pertaining to mathematics, though algebra, geometry, and number were more heavily assessed (Blömeke & Delaney, 2012).

Some studies of teachers' affect in regard to statistics and statistics teaching have used collections of questions from larger scales intended for a broad population. Estrada and Batanero (2008), for example, used a subset of items from the Survey of Attitudes Toward Statistics (Schau, Stevens, Dauphine, & del Vecchio, 1995) that had previously yielded lower scores for teachers. More recently, Harrell-Williams et al. (2014) designed a scale to assess teachers' self-efficacy to teach statistics. It measures teachers' feelings of preparedness to teach content from the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) report for grades pre-K–12 (Franklin et al., 2007). This sort of assessment, which is specifically designed and psychometrically tested to measure affect in regard to *teaching* statistics, is relatively rare.

Some written assessments deal with both cognition and affect related to teaching statistics. Watson (2001) designed a survey to assess teachers' pedagogical content knowledge, self-efficacy to teach statistics, and beliefs about the value of statistics. Watson, Callingham, and Donne (2008) built on this survey to devise a 12-item scale of pedagogical content knowledge. The Statistics Teaching Inventory (Zieffler, Park, Garfield, del Mas, & Bjornsdottir, 2012) contains questions about statistics teachers' teaching practices, course characteristics, assessment practice, teaching beliefs, and assessment beliefs. González (2014) designed a written assessment of teachers' subject matter knowledge, pedagogical content knowledge, and beliefs and conceptions of variability. Instruments that assess aspects of both cognition and affect have the potential to help researchers understand complex relationships among teacher characteristics such as knowledge, beliefs, attitudes, goals, and teaching practices.

10.4.2 Interviews

Clinical interviews allow a high degree of interactivity between the researcher and study participants. They are more time intensive than written assessments. Interviews come in a variety of forms. They may be driven by a formal protocol, such as the StatSmart teacher interview protocol (Watson & Nathan, 2010), which probes the nature of teachers' subject matter knowledge and pedagogical content knowledge. More often, however, interview tasks and questions are designed to meet the specific objectives of a research study. For example, Noll (2011) interviewed graduate assistants to assess their statistical content knowledge of sampling. Participants were asked about written items they had completed and were given some new tasks to solve. Similarly, Browning, Goss, and Smith (2014) conducted interviews with preservice teachers to gain better understanding of the thinking they employed while solving written statistical tasks. Other studies incorporating interviews have probed subjects such as teachers' classroom practices (Casey, 2010), beliefs about the nature of statistics (Leavy et al., 2013), and perceptions of professional development sessions (Peters, Watkins, & Bennett, 2014).

10.4.3 Observations

Written assessments and interviews provide proxies of teachers' classroom practices and quality of instruction, but observation techniques allow researchers to see these firsthand. Using structured observations, researchers can infer the nature of teachers' SKT and knowledge of statistical investigations. For instance, Burgess (2011) illustrated how an SKT framework can guide such observations of teachers' practice. Casey (2010) described an observation-based process for assessing knowledge for teaching statistical association. Pfannkuch (2006) used observations to describe a teacher's knowledge of comparing distributions with box plots. Jacobbe and Horton (2010) used observations of teachers' lessons to detect their misconceptions related to data displays.

Another strategy is to observe teachers' interactions during professional development sessions. Although these are not always firsthand observations of teaching, they still provide data that may not be easily obtained through interviews or written tasks. Wilson et al. (2011) studied video of teacher education sessions to better understand teachers' pedagogical content knowledge in connection with the use of dynamic statistics software. Peters et al. (2014) observed how teachers' learning of measures of center developed as they interacted with one another. Leavy (2010, 2015) observed prospective teachers involved in lesson study. Lesson study engages a group of teachers in planning a lesson, carrying it out, observing students' reactions, and then debriefing on the lesson's effectiveness. Observations of these activities provide a window into teachers' thinking about planning and analyzing lessons. In general, teachers' discourse with one another during professional development can help explain the nature and origin of the knowledge and beliefs that guide their instruction.

Observations of teachers' lessons and professional development sessions typically yield a variety of artifacts. These may include written lesson plans (Garfield & Ben-Zvi, 2008), statistical tasks a teacher assigns to a class (Burgess, 2011), field notes (Casey, 2010), and teachers' responses to professional development tasks (Wilson et al., 2011). Artifacts of this nature can be synthesized with other data to help researchers develop portraits of teachers' cognitive and affective characteristics related to teaching statistics.

10.5 Research on Teachers' Statistical Knowledge

The research models, constructs, and techniques described up to this point in the chapter have been used in studies spanning various statistical content areas. In this section, we summarize findings from two broad, interrelated bodies of literature about teachers' knowledge related to data (Sect. 10.5.1) and uncertainty and statistical inference (Sect. 10.5.2).

10.5.1 Research on Data

We present a brief overview of research on teachers' subject matter knowledge and pedagogical content knowledge in regard to data displays, distribution and variability, associations between variables, and statistical literacy, reasoning, and thinking.

10.5.1.1 Data Displays

Studies of preservice or practicing teachers' graph reading and interpretation skills show a tendency to underestimate the complexities of learning and teaching graphical representations (Leavy, 2010). Teachers tend to express confidence in their understandings of graphical representations and to feel better equipped to teach this topic compared to other statistical ideas (González & Pinto, 2008; Watson, 2001). However, despite their positive attitudes and confidence toward teaching graphs, many educators have limited subject matter knowledge of graphical representations (González & Pinto, 2008; Jacobbe & Horton, 2010; Pierce & Chick, 2013; Sorto, 2004). They sometimes confuse histograms with bar diagrams (Bruno & Espinel, 2009), fail to integrate graphical knowledge with problem context (Burgess, 2002), and have trouble with graph selection and understandings of data type (Leavy, 2010). Monteiro and Ainley (2007) introduced the idea of "critical sense" as a key skill in the analysis and interpretation of graphical artifacts. They investigated critical sense in graphing among prospective primary school teachers from Brazil and England. They found that many preservice teachers did not have adequate mathematical knowledge to read graphs from the daily press. However, the majority displayed an ability to think critically and justify their ideas by combining statistical knowledge with personal experience and contextual knowledge.

A small number of studies have examined teachers' pedagogical content knowledge of graphs. González and Pinto (2008) concluded that teachers need more knowledge of the process of learning statistical graphs and the difficulties that students might have with them. Espinel, Bruno, and Plasencia (2008) observed lack of coherence between prospective primary teachers' graph building and their evaluation of tasks carried out by fictitious future students. Heaton and Mickelson (2002) observed that graph construction often became the endpoint of statistical investigation for preservice elementary teachers, who focused on the technical aspects of graph construction rather than on engaging children in reasoning with the data. However, some studies indicate that using dynamic data exploration tools (e.g., Finzer, 2002; Konold & Miller, 2011) can help teachers portray graph production as a means for understanding data rather than an end in itself (Meletiou-Mavrotheris, Mavrotheris, & Paparistodemou, 2011).

10.5.1.2 Distribution and Variability

Much of the research investigating teachers' reasoning with distributions has focused on their understanding of measures of central tendency, measures of variability, distributional thinking, and procedural aspects of statistics. Pedagogical content knowledge has also been explored.

Studies examining teachers' conceptions of measures of center have focused on the arithmetic mean (e.g., Batanero, Godino, & Navas, 1997; Gfeller, Niess, & Lederman, 1999); the mean, median, and mode (e.g., Groth & Bergner, 2006; Jacobbe, 2012); and the general concept of average (e.g., Begg & Edwards, 1999; Estrada, Batanero, & Fortuny, 2004; Leavy & O'Loughlin, 2006). Evidence from such studies illustrates that attaining deep understanding of these statistical concepts is nontrivial. Like students, many teachers struggle to view measures of central tendency as representative (or "typical") values. Although teachers can readily calculate the mean, they tend not to use it to compare groups (Canada, 2004; Hammerman & Rubin, 2004; Leavy & O'Loughlin, 2006; Makar, 2004; Makar & Confrey, 2002, 2004; Peters, 2009). Like students, teachers may rely upon procedural algorithms and need conceptual understanding (Gfeller et al., 1999; Leavy & O'Loughlin, 2006; Peters et al., 2014; Sorto, 2004).

As with measures of center, teachers' understandings of standard deviation and other formal measures of variation tend to be procedural (Leavy, 2006; Makar & Confrey, 2005; Sorto, 2004). Research indicates difficulties with the concept of variability for teachers of various grade levels (Mooney, Duni, van Meenen, & Langrall, 2014; Vermette, 2013) and similar misunderstandings to those seen in students (e.g., perceiving the normality shape of a distribution as an indication of low variability). Teachers often hold competing beliefs about random variation when the setting of a problem changes (Canada, 2004; Makar, 2004). Kuntze (2014) found that some secondary teachers did not consider learning about statistical variation to be an important instructional goal, though others did recognize the importance of teaching the concept.

Measures of center and measures of spread are inseparable. Conceptual understanding of standard deviation, for example, requires "a dynamic conception of distribution that coordinates changes to the relative density of values about the mean with their deviation from the mean" (Peters, 2009, p. 21). Teachers often have difficulty coordinating understandings of central tendency and dispersion (Dabos, 2014; Lee & Lee, 2011). Many teachers tend to focus either only on the center of the distribution, or on its range, or on small clusters or individual points, rather than integrating different aspects of data distributions (Canada, 2008; Makar & Confrey, 2005; Mooney et al., 2014). When beginning to reason about distributions, teachers can be encouraged to use informal terminology to describe spread and distribution, such as "clump," "bump," "bulk of this data," "scattered," and "bunched" (Canada, 2004; Makar, 2004; Makar & Confrey, 2005). Since children use similar language (e.g., Bakker & Gravemeijer, 2004), Makar and Confrey (2005) suggested recognizing and valuing this informal "variation talk" as a way to encourage intuitive statistical sensemaking.

Explorations with dynamic data software can also help improve teachers' distributional reasoning and pedagogical content knowledge (Canada, 2004; Hammerman & Rubin, 2004; Leavy, 2006; Lee & Lee, 2011; Makar, 2004; Meletiou-Mavrotheris, Paparistodemou, & Stylianou, 2009; Peters et al., 2014). For example, Meletiou-Mavrotheris and Serradó (2012) reported on EarlyStatistics, an intercultural professional development course in which teachers took part in authentic educational activities. The activities gave them the opportunity to reflect on the "big ideas" of statistics and their applications and to explore ways of improving statistics instruction through the adoption of a coherent technology-rich curriculum based on the statistical problem-solving process. Findings from the study indicated that EarlyStatistics met its objectives, improving teachers' knowledge of key statistical ideas including distributions. Data obtained from follow-up teaching interventions in some of the participants' classrooms suggested positive gains in student learning and attitudes toward statistics (Meletiou-Mavrotheris et al., 2011).

10.5.1.3 Associations Between Variables

Some researchers have designed instruction to help teachers confront their potential misunderstandings of association and those of their students. Batanero, Estepa, and Godino (1997) examined whether a computer-based teaching experiment would improve preservice teachers' understanding of association. They found improvement in covariational strategies and reduction in deterministic concepts of association. However, they also found that most teachers retained the belief that a strong association between two variables is adequate for drawing conclusions about cause and effect. Casey (2010) observed three experienced secondary teachers as they taught statistical association and interviewed them immediately following each observation. The research showed that to meet the demands of teaching, the teachers needed substantial knowledge of the concepts' underlying statistical association. For example, they needed to know not only how to compute the value of a correlation coefficient but also why it was computed as it was and the implications of the computation. Teachers also need to understand the nature of informal lines of best fit and criteria for placing them on graphs. Casey and Wasserman (2015) found that teachers hold a variety of conceptions of informal lines of best fit and how they should be placed. Despite the different conceptions, teachers positioned informal best fit lines in approximately the same place the least-squares regression line would appear in a scatterplot.

Along with subject matter knowledge of association, teachers need pedagogical content knowledge. Casey (2014) synthesized the results of three research studies centered on the teaching and learning of linear regression to describe the knowledge needed by teachers regarding learners' conceptions of linear regression. The synthesis illustrated that the knowledge needed to teach linear regression differs in significant ways from the knowledge needed to teach linear functions. Quintas, Ferreira, and Oliveira (2014) compared and contrasted the pedagogical content knowledge of two experienced secondary mathematics teachers as they taught bivariate data. The teachers had difficulty helping students reason about bivariate relationships. It was challenging for them to teach aspects of structure and strength, model fitting, and the role of the linear regression model in predicting events. Both teachers exhibited some of the common misunderstandings and errors with regard to bivariate relationships identified in the literature (e.g., Engel & Sedlmeier, 2011). Such findings indicate a need to design professional development that strengthens teachers' content knowledge and pedagogical content knowledge for teaching association in tandem.

10.5.1.4 Statistical Literacy, Reasoning, and Thinking

The development of students' statistical literacy has become an overarching goal of statistics education internationally. This broadening of the curriculum to encompass statistical literacy, reasoning, and thinking has put considerable demands on teachers (Hannigan et al., 2013). In particular, they must design lessons with engaging contexts (Chick & Pierce, 2008), focus on conceptual understanding (Watson, 2001), and pose critical questions (Reston, Jala, & Edullantes, 2006).

Research sheds light upon factors that influence the design and implementation of instruction that fosters students' statistical literacy, reasoning, and thinking. Burgess (2011) found that the students of a teacher with well-developed SKT were able to progress further with statistical investigations than students of a teacher whose knowledge was less developed. Callingham and Burgess (2014) conjectured that the national curriculum under which teachers operate may influence their approach to teaching statistics, since Australian teachers in their study tended to focus more on procedural aspects of instruction than did their counterparts from New Zealand. Makar and Fielding-Wells (2011) found that challenges in teaching statistical inquiry may stem from difficulties coping with the uncertainties of inquiry, managing classroom logistics, and developing the requisite content knowledge. Mickelson and Heaton (2004) found that the ability to translate content knowledge into effective teaching practices is complex and urged researchers to team with classroom teachers in order to help design meaningful experiences for students. Indeed, during the past decade, several researchers have been experimenting with new innovative models of preservice and in-service teacher training that are focused on inquiry-based instruction and on statistical problem-solving (e.g., Garfield & Ben-Zvi, 2008; Groth, Bergner, Burgess, Austin, & Holdai, 2016; Makar & Fielding-Wells, 2011; Meletiou-Mavrotheris & Serradó, 2012; Serradó, Meletiou-Mavrotheris, & Paparistodemou, 2014).

10.5.2 Research on Uncertainty and Statistical Inference

Uncertainty and statistical inference are challenging ideas for teachers, just as they are for the general population. Researchers have documented teachers' understanding of theoretical probability (Batanero & Díaz, 2012; Watson, 2001), empirical probability (Dollard, 2011; Groth, 2010; Theis & Savard, 2010), informal inference (Canada, 2008; Pfannkuch, 2006), samples and sampling distributions (Green, 2010; Green & Blankenship, 2014; Groth & Bergner, 2005; Maxara & Biehler, 2010; Noll, 2011), and formal inference (Liu & Thompson, 2009; Thompson & Liu, 2005).

Sound reasoning about uncertainty and inference require a departure from deterministic modes of thinking. Such modes of thinking influence teaching practices. Serradó, Azcárate, and Cardeñoso (2006), for example, found that deterministic beliefs about the nature of statistics prevented some teachers from embracing curricular goals related to probability and uncertainty. Liu and Thompson (2009) found that the majority of the high school teachers in their study tended to think deterministically. This made it difficult for the teachers to understand and portray hypothesis testing as a tool for drawing inferences.

Research suggests that enhancing teachers' subject matter knowledge about uncertainty and statistical inference must be given high priority. Building teachers' knowledge of pedagogical structures and tools by itself is not sufficient. Lee and Mojica (2008), for example, found that although a group of teachers involved their students in authentic statistical inquiry that included use of simulation tools, they missed the chance to develop students' understanding of the frequentist approach to probability because of limited subject matter knowledge. Deep understanding of probability is also needed for identifying student errors and implementing effective teaching practices (Maher & Muir, 2014; Paparistodemou, Potari, & Pitta, 2006). Such understanding can be developed through well-designed professional development. For example, Theis and Savard (2010) helped high school teachers design and implement a technology-based instructional intervention. They found that the use of simulation software within the intervention allowed teachers to adopt more inquiry-oriented strategies and begin to incorporate frequentist probability.

Although having subject matter knowledge is necessary for effective teaching of uncertainty, it is not sufficient. Leavy (2010) worked with a group of prospective teachers who demonstrated relatively strong subject matter knowledge about informal inference. However, they had trouble using this knowledge to develop pedagogical contexts for advancing children's learning. In particular, they had difficulty choosing sufficiently complex data, creating engaging contexts, handling unexpected student responses, and scaffolding children's thought processes. In other studies, gaps in pedagogical content knowledge have been framed as contributing factors to teachers' failure to emphasize important probability concepts when writing lesson plans (Chick & Pierce, 2008) and designing productive learning environments for students (Groth, 2010). Accordingly, researchers have begun to develop techniques capable of assessing both subject matter knowledge and pedagogical content knowledge (Meletiou-Mavrotheris, Kleanthous, & Paparistodemou, 2014) and to monitor and adjust their professional development efforts to ensure that they facilitate teachers' development of both of these aspects of SKT (Lee & Hollebrands, 2008; Serradó et al., 2014).

10.6 Teacher Education Frontiers

As illustrated in this chapter and in other reports on statistics teacher education (Franklin et al., 2015), developing cognitive and affective characteristics related to teaching statistics is a complex process deserving concentrated research attention. Such research attention is particularly important in light of the move away from traditional methods of teaching statistics (Batanero & Díaz, 2012). In order to better fulfill teachers' needs in a reform-oriented context, alternative approaches to teacher education have become prevalent. Some alternative approaches are situated within: the context of teachers' classroom practice, deep exploration of statistical content, and technological environments. To conclude the chapter, we summarize some of the work done using these approaches. We do so to encourage others to continue to develop and extend the approaches. We also foreshadow Chap. 12, which describes approaches to professional development in greater detail.

10.6.1 Learning from Teaching Practice

In practice-based approaches, teachers use real classrooms as sites for investigation. Lesson study is one example. Leavy (2010, 2015) used lesson study to develop prospective teachers' knowledge and ability to teach informal inferential reasoning and data handling. Roback, Chance, Legler, and Moore (2006) used lesson study to refine their own approaches to teaching inference. Other practice-based approaches involve researchers collaborating with teachers to design, implement, and analyze instruction. For example, Noll and Shaughnessy (2012) reported on a project in which teachers teamed with five university researchers to design and co-teach lessons to investigate secondary students' conceptions of variability. They found the project to be mutually beneficial; teachers and researchers learned from one another during collaboration. Groth et al. (2016) collaborated with prospective teachers to design instruction suitable for meeting students' learning needs in regard to measures of center and involved the prospective teachers in the process of disseminating the results (Groth, Kent, & Hitch, 2015). Under such approaches, the line between teachers and researchers is intentionally blurred in order to engage teachers in some of the same types of systematic classroom inquiry that are characteristic of formal research.

10.6.2 Immersion in Statistical Content

Examples of approaches that immerse teachers in deep exploration of statistics content can be found at the primary, secondary, and tertiary levels. Reston (2012) explored in-service elementary teachers' conceptions of probability, finding that problem-based learning, inquiry, and statistical investigations promoted stronger conceptual understandings of probability and enhanced pedagogical skills. Makar and Confrey (2002) immersed secondary teachers in focused investigations about student data and studied their statistical reasoning when comparing two groups. They concluded that such an immersion model can help improve teachers' conceptual understanding of inference, their instructional practices, and their disposition toward inquiry. Bargagliotti et al. (2014) developed materials capable of promoting secondary teachers' deep immersion in the study of variability and regression. At the tertiary level, Green and Blankenship (2014) designed a course to develop teaching assistants (TAs) as statistics educators. The course focused on how TAs can foster critical thinking and enhance learning in their classrooms. The TAs left the course with improved conceptual understanding of sampling distributions and strategies for teaching and assessing students.

10.6.3 Technological Environments

Numerous studies have investigated the use of dynamic statistical packages (Finzer, 2002; Konold & Miller, 2011) to develop teachers' knowledge for teaching concepts such as sampling distributions, the central limit theorem, confidence intervals, and hypothesis testing (e.g., Garfield & Everson, 2009; Maxara & Biehler, 2010; Meletiou-Mavrotheris et al., 2014). Such studies indicate that teachers' experimentation with statistical ideas through investigations of authentic and computer-simulated data can help them develop informal inferential reasoning, construct more sophisticated understandings of the logic of inferential statistics, and improve their repertoire of teaching strategies related to statistical inference. Those interested in exploring the potential of dynamic statistics software for supporting teachers' learning can take advantage of resources such as Lee and Hollebrands' (2008) teacher education curriculum that incorporates the software and Madden's (2011) framework describing the characteristics of statistically, contextually, and/or technologically provocative tasks.

Along with dynamic statistics software environments, there are many other technological frontiers to continue to explore for teacher education. These include online communities, mobile devices, and the use of big data in relation to assessment and instruction. Environments and tools of this nature help break traditional boundaries of time, location, and extent of teacher learning (Koehler & Mishra, 2008). As relatively new, emerging technologies, much of the story of their impact on statistics teachers' learning remains to be written.

10.7 Conclusion

Research on cognition and affect related to teaching statistics is a relatively new endeavor. Several opportunities for future research are identified in this chapter. Work remains to be done to more clearly define the elements of SKT, their relationships among one another, and the mechanisms through which they develop. As this work is carried out, it will be important to reconcile and refine different SKT models through systematic academic discourse among researchers working from diverse inquiry paradigms. We also need better understanding of interactions among affect, cognition, and teaching practices. We know that teachers' goals, beliefs, and attitudes influence teaching practices to an extent. The precise nature of the types of goals, beliefs, and attitudes that are most relevant and their degree of impact need further investigation. Additionally, we need better understanding of teachers' knowledge of the impact of social and environmental factors on students' achievement and interest in statistics and their levels of preparedness to help diverse populations of students learn statistics.

Qualitative and quantitative approaches each have roles to play as research on cognition and affect for teaching statistics continues to develop. Some psychometrically and theoretically sound quantitative instruments specific to teaching statistics exist, but many studies still have to rely on instruments developed for the general population. Qualitative research can help define the salient cognitive and affective characteristics to be assessed and can provide vivid portraits of how such characteristics may develop under different circumstances. As this work occurs, we can gain progressively better understanding of teachers' mediating role between statistical content and students.

References

Artigue, M. (2015). Perspectives on design research: The case of didactical engineering. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 467–496). Dordrecht, The Netherlands: Springer.

Ashton, P. T. (1985). Motivation and the teachers' sense of efficacy. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol. 2, pp. 141–171). Orlando, FL: Academic Press.

Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In I. D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer.

Bakker, A., & van Eerde, D. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 429–466). Dordrecht, The Netherlands: Springer.

Ball, D. L., & Bass, H. (2009). *With an eye on the mathematical horizon: Knowing mathematics for teaching to learners' mathematical futures*. Paper Presented at the 43rd Jahrestagung der Gesellschaft für Didaktik der Mathematik, Oldenburg, Germany.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389–407.

Bargagliotti, A., Anderson, C., Casey, S., Everson, M., Franklin, C., Gould, R., et al. (2014). Project-SET materials for the teaching and learning of sampling variability and regression. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Batanero, C., & Díaz, D. (2012). Training school teachers to teach probability: Reflections and challenges. *Chilean Journal of Statistics, 3*(1), 3–13.

Batanero, C., Estepa, A., & Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer-based teaching environment. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE Round Table Conference* (pp. 191–205). Voorburg, The Netherlands: International Statistical Institute.

Batanero, C., Godino, J., & Navas, F. (1997). Concepciones de maestros de primaria en formción sobre promedios (Primary school teachers' conceptions on averages). In H. Salmerón (Ed.), *Actas de las VII Jornadas LOGSE: Evaluación Educativa* (pp. 310–340). Granada, Spain: University of Granada.

Begg, A., & Edwards, R. (1999). Teachers' ideas about teaching statistics. In *Proceedings of the 1999 combined Conference of the Australian Association for Research in Education and the New Zealand Association for Research in Education*. Melbourne: Australian Association for Research in Education & New Zealand Association for Research in Education. Retrieved from http://www.aare.edu.au/data/publications/1999/beg99082.pdf

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1-2), 127–155.

Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM, 44*(3), 223–247.

Brown, S. A., Pitvorec, K., Ditto, C., & Kelso, C. R. (2009). Reconceiving fidelity of implementation: An investigation of elementary whole-number lessons. *Journal for Research in Mathematics Education, 40*(4), 363–395.

Browning, C., Goss, J., & Smith, D. (2014). Statistical knowledge for teaching: Elementary preservice teachers. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Bruno, A., & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training. *International Journal of Mathematical Education in Science and Technology, 40*(4), 473–493.

Burgess, T. (2002). Investigating the "data sense" of preservice teachers. In B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching Statistics*. Cape Town, South Africa: International Association for Statistics Education. Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/1/6e4_burg.pdf

Burgess, T. A. (2011). Teacher knowledge of and for statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 259–270). Dordrecht, The Netherlands: Springer.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics—Challenges for teaching and teacher education* (pp. 57–69). Dordrecht, The Netherlands: Springer.

Callingham, R., & Burgess, T. (2014). How the curriculum shapes teachers' thinking: A comparison of New Zealand and Australian teachers' thinking about statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Callingham, R., Carmichael, C., & Watson, J. M. (2016). Explaining student achievement: The influence of teachers' pedagogical content knowledge in statistics. *International Journal of Science and Mathematics Education, 14*(7), 1339–1357.

Canada, D. (2004). *Elementary preservice teachers' conceptions of variation*. Unpublished doctoral dissertation, Portland State University, Portland, OR.

Canada, D. L. (2008). Conceptions of distribution held by middle school students and preservice teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

Casey, S. A. (2010). Subject matter knowledge for teaching statistical association. *Statistics Education Research Journal, 9*(2), 50–68. Retrieved from http://iase-web.org/documents/SERJ/SERJ9%282%29_Casey.pdf

Casey, S. A. (2014). Teachers' knowledge of students' conceptions and their development when learning linear regression. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International statistical institute.

Casey, S. A., & Wasserman, N. H. (2015). Teachers' knowledge about informal line of best fit. *Statistics Education Research Journal, 14*(1), 8–35. Retrieved from http://iase-web.org/documents/SERJ/SERJ14(1)_Casey.pdf

Chick, H. L., & Pierce, R. U. (2008). Teaching statistics at the primary school level: Beliefs, affordances, and pedagogical content knowledge. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning, 1*, 5–43.

Dabos, M. (2014). A glimpse of two year college instructors' understanding of variation in histograms. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Dollard, C. (2011). Preservice elementary teachers and the fundamentals of probability. *Statistics Education Research Journal, 10*(2), 27–47. Retrieved from http://iase-web.org/documents/SERJ/SERJ10(2)_Dollard.pdf

Eichler, A. (2007). Individual curricula: Teachers' beliefs concerning stochastics instruction. *International Electronic Journal of Mathematics Education, 2*(3), 208–226. Retrieved from http://www.iejme.com/032007/d6.pdf

Engel, J., & Sedlmeier, P. (2011). Correlation and regression in the training of teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 247–258). New York: Springer.

Espinel, M. C., Bruno, A., & Plasencia, I. (2008). Statistical graphs in the training of teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

Estrada, A., & Batanero, C. (2008). Explaining teachers' attitudes toward statistics. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

Estrada, A., Batanero, C., & Fortuny, J. M. (2004). Un estudio sobre conocimientos de estadística elemental de profesores en formación (Prospective teachers' knowledge on elementary statistics). *Educación matemática, 16*(1), 89–111.

Estrada, A., Batanero, C., Fortuny, J. M., & Díaz, C. (2005). A structural study of future teachers' attitudes toward statistics. In M. Bosch (Ed.), *Proceedings of the 4th Congress of the European Society for research in Mathematics Education* (pp. 508–517). Barcelona: IQS Fundemi.

Estrada, A., Batanero, C., & Lancaster, S. (2011). Teachers' attitudes toward statistics. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 163–174). Dordrecht, The Netherlands: Springer.

Finzer, W. (2002). Fathom dynamic data software (Version 2.1) [Computer software]. Emeryville, CA: Key Curriculum Press.

Forgasz, H. J., & Leder, G. C. (2008). Beliefs about mathematics and mathematics teaching. In P. Sullivan & T. Wood (Eds.), *International handbook of mathematics teacher education, Knowledge and beliefs in mathematics teaching and teaching development* (Vol. 1, pp. 173–192). Rotterdam, Taipei: Sense Publishers.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225–256). Reston, VA: National Council of Teachers of Mathematics.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K–12 curriculum framework*. Alexandria, VA: American Statistical Association.

Franklin, C. A., Bargagliotti, A. E., Case, C. A., Kader, G. D., Scheaffer, R. L., & Spangler, D. A. (2015). *The statistical education of teachers*. Alexandria, VA: American Statistical Association.

Garfield, J., & Ben-Zvi, D. (2008). Preparing school teachers to develop students' statistical reasoning. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

Garfield, J. B., & Everson, M. (2009). Preparing teachers of statistics: A graduate course for future teachers. *Journal of Statistics Education, 17*(2), 223–237. Retrieved from http://www.amstat.org/publications/jse/v17n2/garfield.pdf

Gattuso, L. (2008). Statistics in a mathematical context? In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

Gfeller, M. K., Niess, M. L., & Lederman, N. G. (1999). Preservice teachers' use of multiple representations in solving arithmetic mean problems. *School Science and Mathematics, 99*(5), 250–257.

Godino, J. D., Ortiz, J. J., Roa, R., & Wilhelmi, M. R. (2011). Models for statistical pedagogical knowledge. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 271–282). Dordrecht, The Netherlands: Springer.

González, M. T., & Pinto, J. (2008). Conceptions of four pre-service teachers on graphical representation. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

González, O. (2014). Secondary mathematics teachers' professional competencies for effective teaching of variability-related ideas: A Japanese case study. *Statistique et Enseignment, 5*(1), 31–51.

Goodchild, S., & Sriraman, B. (2012). Revisiting the didactic triangle: From the particular to the general. *ZDM, 44*(5), 581–585.

Green, J. L. (2010). Teaching highs and lows: Exploring university teaching assistants' experiences. *Statistics Education Research Journal, 9*(2), 108–122. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ9(2)_Green.pdf

Green, J. L., & Blankenship, E. E. (2014). Beyond calculations: Fostering conceptual understanding in statistics graduate teaching assistants. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education, 38*(5), 427–437.

Groth, R. E. (2010). Teachers' construction of learning environments for conditional probability and independence. *International Electronic Journal of Mathematics Education, 5*, 32–55. Retrieved from http://mathedujournal.com/dosyalar/IJEM_v5n1_2.pdf

Groth, R. E. (2013). Characterizing key developmental understandings and pedagogically powerful ideas within a statistical knowledge for teaching framework. *Mathematical Thinking and Learning, 15*(2), 121–145.

Groth, R. E. (2014). Prospective teachers' procedural and conceptual knowledge of mean absolute deviation. *Investigations in Mathematics Learning, 6*(3), 51–69.

Groth, R. E., & Bergner, J. A. (2005). Pre-service elementary school teachers' metaphors for the concept of statistical sample. *Statistics Education Research Journal, 4*(2), 27–42. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ4(2)_groth_bergner.pdf

Groth, R. E., & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning, 8*, 37–63.

Groth, R. E., Bergner, J. A., Burgess, C. R., Austin, J. W., & Holdai, V. (2016). Re-imagining education of mathematics teachers through undergraduate research. *Council on Undergraduate Research (CUR) Quarterly, 36*(3), 41–46.

Groth, R. E., Kent, K., & Hitch, E. (2015). Journey to centers in the core. *Mathematics Teaching in the Middle School, 21*, 295–302.

Hammerman, J. K., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17–41. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Hammerman_Rubin.pdf

Hannigan, A., Gill, O., & Leavy, A. M. (2013). An investigation of prospective secondary mathematics teachers' conceptual knowledge of and attitudes towards statistics. *Journal of Mathematics Teacher Education, 16*(6), 427–449.

Harrell-Williams, L., Sorto, M. A., Pierce, R. L., Lesser, L. M., & Murphy, T. J. (2014). Validation of scores from a new measure of pre-service teachers' self-efficacy to teach statistics in the middle grades. *Journal of Psychoeducational Assessment, 32*(1), 40–50.

Heaton, R. M., & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education, 5*(1), 35–59.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400.

Jacobbe, T. (2012). Elementary school teachers' understanding of mean and median. *International Journal of Science and Mathematics Education, 10*, 1143–1161.

Jacobbe, T., & Horton, R. M. (2010). Elementary school teachers' comprehension of data displays. *Statistics Education Research Journal, 9*(1), 27–45. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ9%281%29_Jacobbe_Horton.pdf

Koehler, M. J., & Mishra, P. (2008). Introducing TPCK. In AACTE Committee on Innovation and Technology (Ed.), *Handbook of technological pedagogical content knowledge (TPCK) for educators* (pp. 3–29). New York: Routledge.

Konold, C., & Miller, C. (2011). Tinker plots (Version v2.0) [Computer software]. Emeryville, CA: Key Curriculum Press.

Kuntze, S. (2014). Teachers' views related to goals of the statistics classroom – from global to content-specific. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS 9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Leavy, A., & O'Loughlin, N. (2006). Preservice teachers' understanding of the mean: Moving beyond the arithmetic average. *Journal of Mathematics Teacher Education, 9*, 53–90.

Leavy, A. M. (2006). Using data comparison to support a focus on distribution: Examining preservice teachers' understandings of distribution when engaged in statistical inquiry. *Statistics Education Research Journal, 5*(2), 89–114. Retrieved from http://iase-web.org/documents/SERJ/SERJ5(2)_Leavy.pdf

Leavy, A. M. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning. *Statistics Education Research Journal, 9*(1), 46–67. Retrieved from http://iase-web.org/documents/SERJ/SERJ9(1)_Leavy.pdf

Leavy, A. M. (2015). Looking at practice: Revealing the knowledge demands of teaching data handling in the primary classroom. *Mathematics Education Research Journal, 27*(3), 283–309.

Leavy, A. M., Hannigan, A., & Fitzmaurice, O. (2013). If you're doubting yourself, what's the fun in that? An exploration of why prospective secondary mathematics teachers perceive statistics as difficult. *Journal of Statistics Education, 21*(3), n3. Retrieved from www.amstat.org/publications/jse/v21n3/leavy.pdf

Lee, H., & Hollebrands, K. (2008). Preparing to teach mathematics with technology: An integrated approach to developing technological pedagogical content knowledge. *Contemporary Issues in Technology and Teacher Education, 8*(4), 326–341. Retrieved from http://www.citejournal.org/vol8/iss4/mathematics/article1.cfm

Lee, H. S., & Hollebrands, K. F. (2011). Characterising and developing teachers' knowledge for teaching statistics with technology. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 359–369). Dordrecht, The Netherlands: Springer.

Lee, H. S., Kersaint, G., Harper, S. R., Driskell, S. O., Jones, D. L., Leatham, K. R., et al. (2014). Teachers' use of transnumeration in solving statistical tasks with dynamic statistical software. *Statistics Education Research Journal, 13*(1), 25–52. Retrieved from http://iase-web.org/documents/SERJ/SERJ13%281%29_Lee.pdf

Lee, H. S., & Lee, J. T. (2011). Enhancing prospective teachers' coordination of center and spread: A window into teacher education material development. *The Mathematics Educator, 21*(1), 33–47.

Lee, H. S., & Mojica, G. S. (2008). Examining how teachers' practices support statistical investigations. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

Liu, Y., & Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies, 4*(2), 126–138.

Madden, S. R. (2011). Statistically, technologically, and contextually provocative tasks: Supporting teachers' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1-2), 109–131.

Maher, N., & Muir, T. (2014). "I don't really understand probability at all": Final year pre-service teachers' understanding of probability. In J. Anderson, M. Cavanagh, & A. Prescott (Eds.), *Curriculum in focus: Research guided practice. Proceedings of the 37th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 437–444). Sydney: MERGA.

Makar, K. (2004). *Developing statistical inquiry: Prospective secondary mathematics and science teachers' investigations of equity and fairness through analysis of accountability data*. Unpublished doctoral dissertation, University of Texas at Austin, Austin.

Makar, K., & Confrey, J. (2002). Comparing two distributions: Investigating secondary teachers' statistical thinking. In B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching Statistics [CD-ROM]*. International Statistical Institute and International Commission on Mathematical Instruction: Voorburg, The Netherlands.

Makar, K., & Confrey, J. (2004). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 353–374). Boston: Kluwer Academic.

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27–54. Retrieved from http://iase-web.org/documents/SERJ/SERJ4(1)_Makar_Confrey.pdf

Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics:*

Challenges for teaching and teacher education (pp. 347–358). Dordrecht, The Netherlands: Springer.

Martins, J. A., Nascimento, M. M., & Estrada, A. (2012). Looking back over their shoulders: A qualitative analysis of Portuguese teachers' attitudes towards statistics. *Statistics Education Research Journal, 11*(2), 26–44. Retrieved from http://iase-web.org/documents/SERJ/SERJ11(2)_Martins.pdf

Maxara, C., & Biehler, R. (2010). Students' understanding and reasoning with sample size and the law of large numbers after a computer-intensive introductory course on stochastics. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence based society. Proceedings of the 8th International Conference on Teaching Statistics, Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute.

McLeod, D. B. (1992). Research on affect in mathematics: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). New York: MacMillan.

Meletiou-Mavrotheris, M., Kleanthous, I., & Paparistodemou, E. (2014). Developing pre-service teachers' technological pedagogical content knowledge (TPACK) of sampling. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Meletiou-Mavrotheris, M., Mavrotheris, E., & Paparistodemou, E. (2011). Distance learning for teacher professional development in statistics education. *Teaching Statistics, 33*(1), 2–8.

Meletiou-Mavrotheris, M., Paparistodemou, E., & Stylianou, D. (2009). Enhancing statistics instruction in elementary schools: Integrating technology in professional development. *The Montana Mathematics Enthusiast, 16*(1&2), 57–78.

Meletiou-Mavrotheris, M., & Serradó, A. (2012). Distance training of mathematics teachers: The EarlyStatistics experience. *Universities and Knowledge Society Journal (RUSC), 9*(1), 340–353.

Mickelson, W., & Heaton, R. (2004). Primary teachers' statistical reasoning with data. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 327–352). Dordrecht, The Netherlands: Kluwer.

Monteiro, C., & Ainley, J. (2007). Investigating the interpretation of media graphs among student teachers. *International Electronic Journal of Mathematics Education, 2*(3), 188–207. Retrieved from http://www.mathedujournal.com/dosyalar/IJEM_v2n3_5.pdf

Mooney, E., Duni, D., van Meenen, E., & Langrall, C. (2014). Preservice teachers' awareness of variability. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Moore, D. S. (1988). Should mathematicians teach statistics? *The College Mathematics Journal, 19*(1), 3–7.

Nasser, F. M. (2004). Structural model of the effects of cognitive and affective factors on the achievement of Arabic-speaking prospective teachers in introductory statistics. *Journal of Statistics Education, 12*(1), 1–28. Retrieved from http://www.amstat.org/publications/jse/v12n1/nasser.html

National Governors Association for Best Practices & Council of Chief State School Officers. (2010). Common core state standards for mathematics. Retrieved from http://www.corestandards.org/

Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys assessing students' attitudes toward statistics: A systematic review of validity and reliability. *Statistics Education Research Journal, 11*(2), 103–123. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ11(2)_Nolan.pdf

Noll, J., & Shaughnessy, J. M. (2012). Aspects of students' reasoning with variation in empirical sampling distributions. *Journal for Research in Mathematics Education, 43*(5), 509–556.

Noll, J. A. (2011). Graduate teaching assistants' statistical content knowledge of sampling. *Statistics Education Research Journal, 10*(2), 48–74. Retrieved from https://www.stat.auckland.ac.nz/~iase/serj/SERJ10(2)_Noll.pdf

Onwuegbuzie, A. J. (2000). Statistics anxiety and the role of self-perceptions. *Journal of Educational Research, 93*(5), 323–335.

Paparistodemou, E., Potari, D., & Pitta, D. (2006). Prospective teachers' awareness of young children's stochastic activities. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics*. Salvador, Brazil: International Statistical Institute and International Association for Statistical Education.

Peters, S. A. (2009). *Developing an understanding of variation: AP statistics teachers' perceptions and recollections of critical moments*. Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA.

Peters, S. A., Watkins, J. D., & Bennett, V. M. (2014). Middle and high school teachers' transformative learning of center. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal, 5*(2), 27–45. Retrieved from http://iase-web.org/documents/SERJ/SERJ5%282%29_Pfannkuch.pdf

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257–315). Reston, VA: National Council of Teachers of Mathematics.

Pierce, R., & Chick, H. (2013). Workplace statistical literacy for teachers: Interpreting box plots. *Mathematics Education Research Journal, 25*(2), 189–205.

Pierce, R., & Chick, H. L. (2011). Teachers' beliefs about statistics education. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 151–162). Dordrecht, The Netherlands: Springer.

Quintas, S., Ferreira, R. T., & Oliveira, H. (2014). Attending to students' thinking on bivariate statistical data at secondary level: Two teachers' pedagogical content knowledge. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Reston, E. (2012). *Exploring inservice elementary mathematics teachers' conceptions of probability through inductive teaching and learning methods*. Paper presented in the 12th International Congress on Mathematical Education, 8–15 July, 2012, COEX, Seoul, Korea.

Reston, E., Jala, L., & Edullantes, T. (2006). Probing college statistics teachers' instructional goals and classroom practices within a statistical literacy framework. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the 7th International Conference on Teaching Statistics, Salvador, Brazil*. International Association for Statistical Education and International Statistical Institute: Voorburg, The Netherlands.

Roback, P., Chance, B., Legler, J., & Moore, T. (2006). Applying Japanese lesson study principles to an upper-level undergraduate statistics course. *Journal of Statistics Education, 14*(2.) Retrieved from http://www.amstat.org/publications/jse/v14n2/roback.html

Rossman, A., Chance, B., & Medina, E. (2006). Some important comparisons between statistics and mathematics and why teachers should care. In G. F. Burrill & P. C. Elliot (Eds.), *Thinking and reasoning with data and chance: 68th yearbook* (pp. 323–333). Reston, VA: National Council of Teachers of Mathematics.

Saderholm, J., Ronau, R., Brown, E. T., & Collins, G. (2010). Validation of the diagnostic teacher assessment of mathematics and science (DTAMS) instrument. *School Science and Mathematics, 110*(4), 180–192.

Schau, C., Stevens, J., Dauphine, T., & del Vecchio, A. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement, 55*(5), 868–875.

Scheaffer, R. L. (2006). Statistics and mathematics: On making a happy marriage. In G. F. Burrill & P. C. Elliot (Eds.), *Thinking and reasoning with data and chance: 68th yearbook* (pp. 309–321). Reston, VA: National Council of Teachers of Mathematics.

Sedlmeier, P., & Wassner, C. (2008). German mathematics teachers' views on statistics education. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

Serradó, A., Azcárate, P., & Cardeñoso, J. M. (2006). Analyzing teacher resistance to teaching probability in compulsory education. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics*. International Statistical Institute and International Association for Statistical Education: Salvador, Brazil. Retrieved from http://www.stat.auckland.ac.nz/~iase

Serradó, A., Meletiou-Mavrotheris, M., & Paparistodemou, E. (2014). Early statistics: A course for developing teachers' statistics technological and pedagogical content. *Statistique et Enseignement, 5*(1), 5–29.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–23.

Silverman, J., & Thompson, P. W. (2008). Toward a framework for the development of mathematical knowledge for teaching. *Journal of Mathematics Teacher Education, 11*(6), 499–511.

Simon, M. (2006). Key developmental understandings in mathematics: A direction for investigating and establishing learning goals. *Mathematical Thinking and Learning, 8*(4), 359–371.

Sorto, M. A. (2004). *Prospective middle school teachers' knowledge about data analysis and its application to teaching*. Unpublished doctoral dissertation, Michigan State University.

Sullivan, P., & Mousley, J. (2001). Thinking teaching: Seeing mathematics teachers as active decision makers. In F.-L. Lin & T. J. Cooney (Eds.), *Making sense of mathematics teacher education* (pp. 147–163). Dordrecht, The Netherlands: Kluwer.

Tarr, J. E., Reys, R. E., Reys, B. J., Chávez, Ó., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*(3), 247–280.

Theis, L., & Savard, A. (2010). Linking probability to real-world situations: How do teachers make use of the mathematical potential of simulation programs? In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics, Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute.

Thompson, P. W., & Liu, Y. (2005). Understandings of margin of error. In S. Wilson (Ed.), *Proceedings of the 27th Annual Meeting of the International Group for the Psychology of Mathematics Education*. Virginia Polytechnic and State Institute: Vicksburg, VA.

Venkat, H., & Adler, J. (2014). Pedagogical content knowledge in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 477–480). Dordrecht, The Netherlands: Springer.

Vermette, S. (2013). *Le concept de variabilité chez des enseignants de mathématiques au secondaire*. Unpublished doctoral dissertation, Université de Sherbrooke, Sherbrooke.

Wassong, T., & Biehler, R. (2010). A model for teacher knowledge as a basis for online courses for professional development of statistics teachers. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics. Ljubljana, Slovenia*.

Watson, J. M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education, 4*(4), 305–337.

Watson, J. M., Callingham, R., & Donne, J. (2008). Establishing pedagogical content knowledge for teaching statistics. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and* *teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Retrieved from http://www.ugr.es/~icmi/iase_study

Watson, J. M., & Nathan, E. L. (2010). Approaching the borderlands of statistics and mathematics in the classroom: Qualitative analysis engendering an unexpected journey. *Statistics Education Research Journal, 9*(2), 69–87. Retrieved from http://iase-web.org/documents/SERJ/SERJ9%282%29_Watson_Nathan.pdf

Wilson, P. H., Lee, H. S., & Hollebrands, K. F. (2011). Understanding prospective mathematics teachers' processes for making sense of students' work with technology. *Journal for Research in Mathematics Education, 42*(1), 42–67.

Yang, K.-L. (2014). An exploratory study of Taiwanese mathematics teachers' conceptions of school mathematics, school statistics, and their differences. *International Journal of Science and Mathematics Education, 12*(6), 1497–1518.

Zieffler, A., Park, J., Garfield, J., del Mas, R., & Bjornsdottir, A. (2012). The statistics teaching inventory: A survey on statistics teachers' classroom practices and beliefs. *Journal of Statistics Education, 20*(1), 1–29. Retrieved from http://www.amstat.org/publications/jse/v20n1/zieffler.pdf

Part III
Contemporary Issues and Emerging Directions in Research on Learning and Teaching Statistics

Janet Ainley and Dave Pratt

In this part of the handbook the focus is on looking forward to anticipate future directions for research in statistics education. The section contains four survey chapters, exploring theoretical frameworks (Chap. 11), the curriculum (Chap. 12), building capacity in teacher education (Chap. 14), and the design of learning environments (Chap. 16). In each of these the authors have considered issues which are of current concern for statistics education research, and built on these to raise questions which they feel should be addressed in future research. Interspersed with these are two "Reflections" chapters, edited by Rob Gould, which are collections of shorter pieces of writing. In Chaps. 13 and 15 experienced researchers and curriculum developers from a wide range of backgrounds reflect on aspects of statistics education and offer suggestions for the future.

Despite the diversity of focus of the six chapters, some common themes emerge which have important implications both for developments in the learning and teaching of statistics and for the research which can inform those developments. Rather than describing the contents of individual chapters, in this introduction we discuss three themes which we see as particularly significant for statistics education research in the future: the need for an holistic approach to change, the importance of statistical literacy for engaged citizenship, and the vital role of technology.

The need for change to be driven by holistic approaches which engage with the complexity of learning and teaching in real classrooms is echoed in much of the writing in Part III. Approaches which aim to identify and improve individual aspects of pedagogy, such as the introduction of a new digital tool, a change in curriculum structure or assessment procedures, or novel delivery of professional development, have had successes within specific fields, but often very limited impact on sustained improvement. Such initiatives are all too easily overwhelmed by other constraints: a new classroom resource may be overlooked if it does not relate to high stakes assessments; changes in the curriculum can only be effectively implemented if teachers have suitable professional development; a new approach to teaching one concept or technique may be at odds with the overall pedagogic style of the classroom.

There is a very real challenge for research here. Piecemeal changes are generally supported by research approaches that are designed to avoid complexity by focusing on trying to isolate particular aspects of classroom context and measure the effectiveness of making changes to these. Acknowledging the need for holistic change requires research methodologies which engage with complexity, rather than trying to manage it. Nilsson et al. in Chap. 11 discuss design research as a means to sensitize the researcher or a team of researchers to a complex ecology. They explain that such approaches seek to develop new theories, including ontological innovations, that might in the future offer accounts of holistic learning environments as envisaged by Ben Zvi et al. in Chap. 16.

A significant constraint on the development of the learning and teaching of statistics is the status that statistics has in the school curriculum. This varies from country to country, but often statistics is part of the mathematics curriculum, and so competes for time and attention with other areas of mathematics. Ponte and Noll refer in Chap. 14 to the relative lack of confidence which many mathematics teachers have about teaching statistics (as well as how teacher education courses might begin to address that). As a result, statistics is always in danger of being squeezed into corners of the timetable, and thus seen as largely unimportant for most students. Those students who need statistical skills to support their study in science or social science will be given tailored courses at College or University. At school level, statistics may be given the minimal attention needed to prepare students for external examinations.

Perhaps the low level of attention currently given to statistics in schooling contributes to the second theme emerging from the chapters in Part III: the growing need for all citizens to have some level of statistical literacy. Increasingly a wide variety of data are presented in advertising and news media, and drawn on in political debates, leaving those who are unable to adopt a critical perspective open to misleading arguments. New approaches to statistical literacy are beginning to emerge. For example, in the reflections presented in Chap. 15, school students explore personally meaningful issues through data in the public domain, and journalists use socially and politically oriented data to enhance their statistical appreciation.

In Chap. 12, Pfannkuch imagines a much richer curriculum than what is generally currently available to students, immersing students in data in ways that might eventually lead to widespread statistical literacy. In the first set of reflections following Chap. 12, there is discussion about new developments that will support a data-rich society. Pfannkuch recognizes that such a curriculum would need to be taught by teachers who have experienced the training methods proposed by Ponte and Noll in Chap. 14.

Underpinning both the need for holistic approaches and the need for critical statistical literacy is the role of technology, which gives access to data and provides the analytical tools in computer-based environments that support learners of all ages to present, represent, and model data. In the reflections in Chap. 15, there is discussion about how technology supports access to data and the need for students to learn technological skills that give them access to powerful tools.

In summary, we regard the three emerging themes as powerful indicators of where statistics education might seek to grow in the next decade or two.

Chapter 11
The Nature and Use of Theories in Statistics Education

Per Nilsson, Maike Schindler, and Arthur Bakker

There is nothing so practical as a good theory.

(Lewin, 1951, p. 169)

The notion of "theory" is crucial in any scholarly or scientific discipline, including research on the teaching and learning of mathematics.

(Niss, 2007, p. 1308)

Abstract This chapter presents a literature review of theories used to frame and underpin Statistics Education Research. The aim is to describe, characterize and arrange the nature and use of theories in SER and hint at some potential trends and required directions for further theorizing the SER discipline. The review includes empirical research papers, published from 2004 to 2015, and focuses on students' learning of statistics or probability at the primary and secondary school level. The number of papers that fulfilled our inclusion criteria was 35.

Electronic supplementary material: The online version of this article (https://doi.org/10.1007/978-3-319-66195-7_11) contains supplementary material, which is available to authorized users.

P. Nilsson (✉)
School of Science and Technology, Örebro University, Örebro, Sweden
e-mail: Per.Nilsson@oru.se

M. Schindler
Faculty of Human Sciences, University of Cologne, Cologne, Germany
e-mail: maike.schindler@uni-koeln.de

A. Bakker
Freudenthal Institute, Utrecht University, Utrecht, The Netherlands
e-mail: a.bakker4@uu.nl

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_11

We distinguish five types of theories used in SER: *Statistical Product Theories*, *Statistical Process Theories*, *Theories with a Didactical Focus*, *Theories in Mathematics/Science Education* and *Theories with a Broader Range on Epistemological Aspects*. For further theoretical elaboration, we argue that SER pay attention to the relationship between personal and formal views of statistics, to the dynamics between categories or levels in student thinking and to the role of technology and context in the learning of statistics and probability. We end the chapter by thinking through potential benefits of a semantic theory, *inferentialism*, that has been proposed as underpinning research on statistical inference.

Keywords Theory • Scientific quality • Literature review • Inferentialism • Empirical studies • Primary and secondary school

11.1 Introduction

This chapter focuses on theories in Statistics Education Research (SER). Theories are crucial to scientific work. The development of the scientific culture of a discipline is critically linked to how the discipline takes into account and contributes to the development of theories (Feuer, Towne, & Shavelson, 2002). By means of theories and theorizing, scientific work aims at bringing order to complex phenomena in order to understand them and explain them and being able to predict their behaviour (Bikner-Ahsbahs & Prediger, 2010; Sriraman & English, 2010). Theories support the researcher to build on others' work, achieve scientific trustworthiness, go beyond common sense, generalize across situations, make valid and reliable interpretations and so on (Bakker & Smit, 2017; Lester, 2010; Silver & Herbst, 2007).

Despite the crucial role of theory for development and quality of scientific research, Lester (2010) raises concerns about a lack of theory and philosophy in the current culture of mathematics education research. In our view, the SER research community needs to take these concerns seriously in order to inform practice and to promote incitements for further development of SER as a scientific discipline.

Taking a meta-perspective on research in SER through the lens of theories is timely for two reasons. Firstly, as disclosed by the different chapters of the present handbook, the main mission of SER is to study and improve the teaching and learning of statistics, which means that SER includes many epistemological concepts and processes that require a theoretical treatment. Without a theoretical elaboration of epistemological and methodological decisions, it becomes difficult to assess the scientific quality of research and to compare and build on research. Secondly, the interest in SER has increased immensely over the past decades, which can be seen in the increased proportion of publications of SER in scientific journals but also increased attention to statistics in curricula in many countries. However, an increased quantity of research does not necessarily imply that the quality of research has increased in the same proportion. There are many factors affecting an increasing proportion of published research. For instance, many universities base their allocation of research funding on the number of publications and citations researchers can present.

As a result, we have a closer look at what this growth actually entails in terms of theories used to frame and underpin SER. As far as we know, no such study has previously been conducted in SER. Theory is of course not a homogeneous category; hence, there is a need to identify the types of theories used in SER, for what purposes theories are used, and to identify potential trends and required directions to strengthen the SER discipline further. To this end, the present chapter presents a literature review of the theoretical work apparent in SER over the past 11 years.

With the focus on theories, we emphasize that the review is not about topics and research findings of SER. Indirectly, of course, topics and findings come into play since theory and theorization in SER are brought to bear when studying some topic of learning or teaching statistics. Moreover, our review is to be considered a configurative literature study (Gough, Oliver, & Thomas, 2013). This means that our ambition is to describe, characterize and arrange (configure) the nature and use of theories in SER and, based on this review, hint at some potential trends and required directions for further theorizing the SER discipline. For reasons of feasibility, we have restricted ourselves to empirical research on teaching and learning at the primary and secondary school level.

Whether research in statistics education uses theories depends of course on what is meant by theory. In this chapter, we explore theory in a broad sense. We are as inclusive as possible in order to account for the nature and use of types of theories emerging in empirical studies on teaching and learning in SER. We begin by grounding the review process in a theoretical consideration of structural features of theories in the nearest educational discipline, that of mathematics education, and what different roles theories are supposed to take in the discipline of mathematics education research. By this discussion, we lay the foundation for our methodological choices of developing and explicating our search strategies and how we approached the configurative analysis and synthesis of our mapping of theories used in SER.

11.2 The Contribution of Theories in Mathematics Education and Other Disciplines

As Groth (2015, p. 4) observed, "Statistics education has begun to mature as a discipline distinct from mathematics education."[1] In many ways, mathematics education research has preceded SER, for example, in terms of domain-specific conferences and journals. SER further deploys many methodological and theoretical approaches that have shown their potential in mathematics education research, and much—but certainly not all—SER is being carried out by researchers with a

[1] We refer the reader to Chap. 2 in this volume for further discussion of the nature and history of statistics education.

background in mathematics education. Hence it makes sense to start with literature on theories in mathematics education before we draw on more general literature.

Although there is no shared definition of theory and theoretical frameworks in mathematics education (Assude, Boero, Herbst, Lerman, & Radford, 2008; Bikner-Ahsbahs & Prediger, 2014), there seems to be some consensus that theory is an organized, consistent and coherent system of terminology, concepts and principles (Sriraman & English, 2010) and claims and predictions about some phenomenon (Niss, 2007). A theory is similarly characterized as a system of concepts and principles that serve as means to explain observed phenomena, to make prediction of yet unobserved phenomena and to guide the researcher in asking questions, formulating hypothesis and identifying key variables and relationships to investigate (Silver & Herbst, 2007).

Mason and Waywood (1996) distinguish background theories and foreground theories in mathematics education. Background theories concern the theoretical foundation a study relies on. The distinction "plays an important role in discerning and defining what kind of objects are to be studied" (p. 1058). To conduct a profound investigation of, for example, students' concept formation in terms of fractions, scholars would have to clarify theoretically what is meant by "concept" and "concept formation." Background theories therefore refer to ontological and epistemological ideas as well as their methodological implications for investigating specific topics (cf. Bikner-Ahsbahs & Prediger, 2014). This is, for example, the case when Vygotsky's (1978) learning theory or Biggs and Collis's (1982) SOLO model is used in SER. Neither theory originates from mathematics or statistics education, but both frame the investigations. By contrast, some theories concern the object of study itself; these are foreground theories. "*Foreground theories* are local theories *in* mathematics education" (Bikner-Ahsbahs & Prediger, 2014, p. 6). Foreground theories are therefore related to the research objective and the research questions. For instance, Watson and Callingham's (2003) hierarchy levels of developing concepts of variation are a foreground theory in Reading's (2004) article that addresses students' descriptions of variation. In the context of SER, foreground theories typically relate to theoretical constructs in which statistical structures, ideas and concepts are salient.

diSessa and Cobb (2004) detail the nature of different theories relevant for research in mathematics education. They distinguish between grand theories, orienting frameworks, frameworks for action and domain-specific instructional theories. Skinner's behaviourist theory and Piaget's system theory of intellectual development are two examples of grand theories. Even if grand theories have a prominent position in educational research, they appear to be too general on their own to provide guidance for explaining and supporting the learning of specific mathematical topics. Orienting frameworks, such as constructivism (von Glasersfeld, 1995) or communities of practice (Lave & Wenger, 1991), provide general support for specifying issues of learning, teaching and instructional design, whereas frameworks for actions concern analytical constructs of a more or less general prescriptive character (diSessa & Cobb, 2004). Domain-specific instructional theories are also of a prescriptive nature as they are typically specific to a domain or even learning trajectory of certain content and the means by which this trajectory can be supported. An ontological innovation is an explanatory construct that cuts across the aforementioned hierarchy and helps to do design work. diSessa and Cobb (2004) illustrated the idea of ontological innovations by referring to their own findings of socio-mathematical norms (Yackel & Cobb, 1996) and meta-representational competence (diSessa & Sherin, 2000). The building blocks of probability literacy (Gal, 2005) and Pratt's (2000) resources related to students' conceptualization of randomness could be considered two representatives of ontological innovations in the field of SER.

diSessa and Cobb's (2004) categorization of theories does not only distinguish theoretical frameworks in terms of analytical level and character. It also points to the different efforts of educational research in using and developing descriptive, explanatory and prescriptive theoretical constructs (McKenney & Reeves, 2012; Silver & Herbst, 2007). The primary goal of descriptive theoretical constructs is to distinguish, narrate and characterize teaching and learning processes. A descriptive construct contains limited explanatory power for why something happens or predictive power for how to act to make something happen. Explanatory theories help research understand practice, such as making research alert to what has created a certain phenomenon and disclosing the causes for why something is happening (Silver & Herbst, 2007). In contrast to descriptive theories, prescriptive theories provide ideas for *how* we can make things happen (McKenney & Reeves, 2012). A framework is prescriptive if it helps in providing advice for instructional planning and actions. Frameworks for actions and domain-specific instructional theories contain elements with a prescriptive nature (Bakker & van Eerde, 2015).

In educational research, it often occurs that descriptive concepts or theories are too easily used for prescriptive purposes (as observed by Säljö, 2003, 2011). We can think of *communities of practice* (Lave & Wenger, 1991), which were originally analytic tools for understanding what happened in naturalistic settings. However, they soon became used as prescriptive tools: Classrooms or even groups of students had to become communities of learners (Brown & Campione, 1994). Without careful consideration of this transition, what is known from descriptive research may be too easily translated to prescription, so projected on how education should be shaped. For SER, let us consider a fictitious example. Konold, Higgins, Russell, and Khalil (2015) have made a famous distinction between ways in which students may view data sets, ranging from data as pointers, case values and classifiers up to aggregates. This is a descriptive categorization. Imagine what could happen if educators were not aware of this. If they slip from a descriptive hierarchy to the idea that statistics education should sequence activities along this hierarchy, they may assume that students have to consider data points as case values before they learn to see them as aggregates. Delineating instructional sequences on this assumption may actually hamper student development. Prescriptive ideas (e.g., the idea of growing samples discussed later) have a different nature: They suggest what could be done to achieve particular learning goals and what student answers can be expected.

The goal of the present chapter is to present a categorization of theories used in SER and to make recommendations for the future about the use of theory in SER. The review addresses the following research questions:

1. What is the nature of theories used in Statistics Education Research?
2. To what extent are different types and combinations of theories used in Statistics Education Research?

With input from the categorization or mapping, we then continue to discuss theoretical issues in SER and suggest possible directions to further theory development in SER. We are not aware of a similar undertaking in mathematics education research. Lerman, Xu, and Tsatsaroni (2002) reviewed 20 years of theory development in one journal, *Educational Studies in Mathematics*, but their study focused on topics, audience and relations of mathematics education research with other disciplines and relations to other official agencies.

11.3 Methods

We pursue these research questions by means of a configurative literature study (Gough et al., 2013) of existing research articles in SER, in which we were able to characterize the nature and use of theories in a systematic and manageable way.

Our review of theories in SER emphasizes the ways researchers ground their research in the introduction and theoretical background sections of their publications. We will not study how theories are actually put into action in the methodological or analytical work of a study. Of course it may be the case that an article reports on a theory in a substantial way in the background but fails in applying the theory later on in the analysis. Also, theories might only appear in the methods and analysis sections without being mentioned in the introduction and theory sections. We are aware that this may be the case and will take this into consideration when reflecting on our results of the mapping.

In order to arrange and characterize the nature and use of theoretical approaches in SER, we focus on empirical research articles. We characterize the landscape of theories in SER by observing and categorizing the theories that are being used as a basis of investigations in empirical research of SER. In a subsequent step of the review, we elaborate on what can be inferred from this landscape.

11.3.1 Search Procedure

In the following, we describe our search process for developing our mapping of theories in SER. For manageability reasons, we restricted our search to a limited set of journals. In the field of SER, we chose:

- *Statistics Education Research Journal* (SERJ)

In the field of mathematics education research, we included the three journals on the Web of Science list (as of 2015):

- *Educational Studies in Mathematics* (ESM)
- *Journal for Research in Mathematics Education* (JRME)
- *Mathematical Thinking and Learning* (MTL)

For being able to see trends of theory use, we decided to include roughly the last decade of publications: 2004 until February 2015. In the aforementioned journals, we searched for empirical articles.

11.3.2 Criteria for Inclusion of Articles

First, we focused on research in statistics education at the *primary and secondary school level*. We decided to exclude research on *tertiary level* (e.g., "Introductory statistics course tertiary students' understanding of p-values" by Reaburn, 2014) or specifically at the *university level* (e.g., "Roles of technology in student learning of university level biostatistics" by Xu, Zhang, Su, Cui, & Qi, 2014). Our main reason for narrowing our focus down to the school years was to get a manageable selection. Here, we build on our experience that research conducted in relation to higher education often rather emphasizes mathematical content in favour of epistemological theories about teaching and learning. However, we are aware that there are certain issues that require specific theories in the post-secondary or higher education programmes, which differ from teaching and learning issues at the primary and secondary school level. Research conducted on vocational education constitutes one such example as it typically deals with phenomena such as boundary crossing between school and work settings that are not prominent in general education (Bakker & Akkerman, 2014), though general education can learn much from vocational and professional practice (Bakker, 2014; Dierdorp, Bakker, van Maanen, & Eijkelhof, 2014).

Second, we focused on research that addresses the *epistemology of statistics education*, with a clear focus on students' learning of statistics or probability. Therefore, we excluded articles that focused on, for example, students' *attitudes* towards statistics (e.g., "Students' attitudes toward statistics across the disciplines: A mixed-methods approach" by Griffith, Adams, Gu, Hart, & Nichols-Whitehead, 2012), even though we acknowledge that the affective side of SER is extremely important and requires more attention. By restricting our focus to students' learning, we also excluded articles that focused exclusively on *teachers* (e.g., "Using an APOS framework to understand teachers' responses to questions on the normal distribution" by Bansilal, 2014).

Third, we limited our search to articles that were published in *English* (e.g., the article by Mary & Gattuso, 2005, was excluded).

11.3.3 Screening Process: Focus on Titles and Abstracts

Based on the aforementioned criteria, we screened the articles in order to select them for the following data analysis. The screening process that led to the selection of articles was conducted in two steps:

In the first step, we applied the aforementioned criteria on the titles of the articles. Here, we paid attention to signal words.

1. Signal words that indicated an *epistemological orientation*.
   Signal words such as "learning," "understanding," "thinking," "awareness," "conceptions" and "conceptual," "reasoning," "inference," "tasks," "activity," "interaction," "strategies," "representations," "teaching," "instruction," etc. indicated an inclusion of the articles.
2. Signal words that indicated the *school level*.
   Signal words such as "university," "workplace," "tertiary," "vocational," etc. led to an exclusion of the articles, unless there were other aspects in the title that led to an inclusion (e.g., when the learning of university students as well as of secondary school students was considered).
3. Signal words that indicated *teachers*
   Signal words such as "teachers" led to an exclusion, unless there were other aspects in the title that led to an inclusion (e.g., when not only the teachers' but also secondary school students' learning was considered).

In case that the information given in the title was not sufficient to conclude on the relevance of the article, in a second step, we screened the *abstracts* for further information regarding the orientation of the article and the objects of study. If necessary, the *methods section* was also taken into account.

11.3.4 Data Analysis: Scanning and Coding the Articles

Based on the screening process, 35 empirical articles were included in the next step of our review procedure: 18 from SERJ, 9 from ESM, 6 from MTL and 2 from JRME.[2]

To answer the *first research question*, we scanned these 35 articles for the theoretical approaches that were used. First and foremost, we focused on the introduction and the theoretical parts of the articles. There were articles that did not label theoretical sections as such. In these cases, we scanned the paragraphs located before the research questions of the articles.

Our theoretical background, presented above (diSessa & Cobb, 2004; Mason & Waywood, 1996; McKenney & Reeves, 2012), was used as an overall, open-minded

[2]An overview of the review of the 35 articles can be found at the Handbook's website in *Springer Link*.

frame for the scanning and categorization of the theories addressed in the articles reviewed. Theories were labelled, whenever they were mentioned in the articles with both a statement and a reference. For instance, say that an article dealt with Vygotsky's theory of understanding as dialogical shared building of meanings. To be included in the map, the article should not only add a reference to the theory e.g.,(Vygotsky, 2001), but also provide an explanation of it. At the end, we discerned four main categories of theories used in SER that focused on teaching and learning of statistics and probability in primary and secondary school. In one of the main categories, on theories of statistics, we could discern two subcategories. The categories we discerned were:

- Theories of statistics (ToS)
  - Statistical Product Theories (SPdT)
  - Statistical Process Theories (SPcT)
- Theories with a Didactical Focus (TDF)
- Theories in Mathematics/Science Education (TMSE)
- Theories with a Broader Range on Epistemological Aspects (TEA)

This categorization was then used to answer our second research question on the extent of different types and combinations of theories used for grounding Statistics Education Research.

11.4 Results

11.4.1 The Nature of Theories Used in SER

In response to our first research question, we start by characterizing the different categories of theories used in empirical SER articles. After a characterization of the categories, we then elaborate on certain phenomena that we found among these categories.

11.4.1.1 Theories of Statistics (ToS)

Theories belonging to two categories concerned the nature of statistics and probability. They can be related to what Mason and Waywood (1996) refer to as foreground theories as they concern the object of study itself and are specific to statistics education. These categories were about the notion of statistical knowledge and what it means to be knowledgeable in statistics. They comprised theories in relation to what students are expected to learn and master regarding the subject. Given the inclusion criteria, all articles of our survey presented some subject matter theory discussion. From our review of subject matter theories, we distinguished two kinds of foreground theories of statistics: *product theories* and *process theories*. In subsequent analysis of these theories, we noted that they are underpinned by either the structure of statistics or probability itself or by empirical results of student understanding or reasoning in relation to the subject matter. We first elaborate the meaning of Statistical Product Theories and Process Theories. We continue the section by elaborating on the underpinnings of these two subcategories.

11.4.1.2 Statistical Product Theories (SPdT)

This category contains theoretical approaches (sometimes referred to as models or frameworks) that address and conceptualize particular statistical concepts or representations. The approaches focus on a single or a limited set of the big ideas of statistics and/or probability, such as variability, average, samples, and graphs (Shaughnessy, 2007), randomness and independence (Gal, 2005), or the role of sample space and comparing probabilities (Nilsson, 2009). Statistical Product Theories relate to what is referred to as central statistical content of many curricula (e.g., National Council of Teachers of Mathematics, 2000; Swedish National Agency for Education, 2012). We use the term “product” to refer to such historically developed content, hence the term “product theories”.

11.4.1.3 Statistical Process Theories (SPcT)

This subcategory covers theoretical approaches (models or frameworks) that focus on conceptualizing and modelling steps and processes involved in statistical investigations. The approaches relate to process standards (National Council of Teachers of Mathematics, 2000; Swedish National Agency for Education, 2012) and deal with statistical knowledge, which emphasizes students becoming engaged in formulating statistical and probabilistic questions, collecting data, analysing data and drawing data-based conclusions and inferences (Paparistodemou & Meletiou-Mavrotheris, 2008). This subcategory can be compared to what Shaughnessy (2007) refers to as models of statistical thinking. We choose to label this category according to the processes involved in statistical work, to clearly separate these approaches from those that focus on teaching and learning specific to some statistical product.

The categories of statistical products and processes revealed to us that their theoretical underpinning took a theoretical *disciplinary perspective* and/or an empirical *student perspective* (Fig. 11.1).

The disciplinary perspective is based on an analysis of the statistics as such and not on empirical data on how students think and reason. This perspective connects to the tradition of *Stoff-didactics* (Steinbring, 2008; Straesser, 2014) and mathematical and historical phenomenology (Freudenthal, 1983) and is intended to describe the mathematical/statistical structure and what it means to be knowledgeable in statistics, based on a normative, discipline-oriented perspective. In short, taking a disciplinary perspective is about theorizing the learning object (Marton, Runesson,

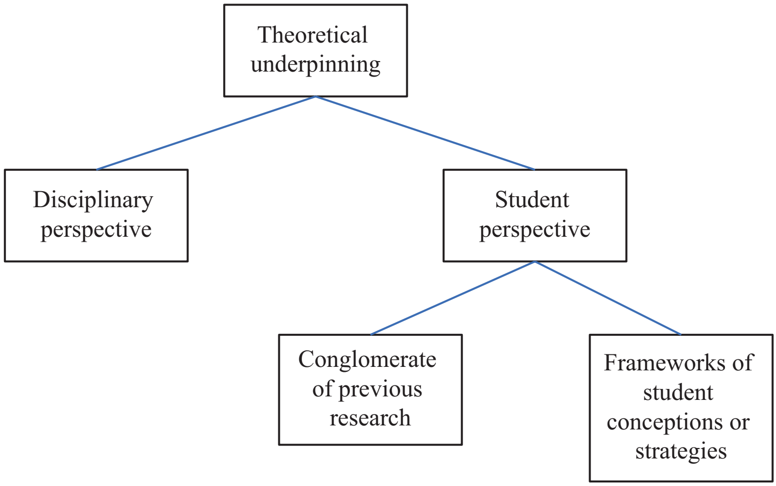


Fig. 11.1 Structure of theoretical underpinning of statistical theories

& Tsui, 2004) in relation to what it is that students should understand or be able to perform in order to count as being knowledgeable in statistics. Freudenthal's idea of a didactical phenomenology then is to think through what the insights from a mathematical and/or historical phenomenology along with empirical educational studies imply for the teaching and learning of the subject matter (see Bakker, 2004; Bakker & Gravemeijer, 2006, for examples in statistics education).

By a student perspective, we mean that the studies referred to research documentation on how students perceive statistical content and express statistical reasoning. In some cases this underpinning was made according to a more or less discreetly outlined conglomerate, a synthesis, of empirical findings from previous research. In other cases, this underpinning took the form of adopting and using structured cognitive frameworks for guiding research. Konold et al.'s (2015) description of four general perspectives that students use in working with data illustrates such a framework.

If we compare studies that conceptualize statistical products with studies that conceptualize statistical processes, we note that the majority of studies focusing on statistical products conceptualize these constructs from a student perspective. The focus is on students' conceptions and, especially, on detecting shortcomings and misconceptions but also emergence and development of statistical reasoning. Definitions of the statistical concepts in question and what it means to understand the concepts are implicit in the studies. In relation to the stage-level structured cognitive frameworks, we observe an implicit mathematical analysis and judgement in that higher levels are considered statistically more sophisticated than the lower levels. However, it is not always clear if the levels of the frameworks are guided and validated by an explicit account of progression from a statistical (content) perspective.

The theoretical underpinning of statistical processes is mainly done by describing processes of statistical work, which is based on the practice developed within the discipline (e.g., Wild & Pfannkuch, 1999). We consider the purpose of this approach being to describe what kind of abilities and understandings statistics education should aim for. Gil and Ben-Zvi (2011) illustrate this perspective in relation to informal inferential statistical reasoning:

> *Informal Inferential Statistical Reasoning* (IIR) refers to the cognitive activities involved in informally drawing conclusions (generalizations) from data (samples) about a wider universe (the population), while attending to the strengths and limitations of the sampling and the drawn inferences (Ben-Zvi, Gil, & Apel, 2007) and "articulating the uncertainty embedded in an inference" (Makar & Rubin, 2009, p. 85). Rubin, Hammerman, and Konold (2006) considered IIR as statistical reasoning that involves consideration of numerous dimensions: properties of data aggregates, the idea of signal and noise, various forms of variability, ideas about sample size and the sampling procedure, representativeness, controlling for bias and tendency (p. 88).

The disciplinary underpinning is often presented in the form of taxonomies or frameworks, for example, Gal's (2005) probabilistic literacy model and Watson's (1997) three-tier hierarchy, phases of statistical investigations (Pfannkuch & Wild's, 2004, PPDAC model) and analytical categories of informal statistical inference (Makar & Rubin, 2009).

11.4.1.4 Theories with a Didactical Focus (TDF)

This category comprises theoretical approaches in which didactical aspects are taken into consideration as means to support learning (instructional activities, computer tools, teaching). Theories in this sense may concern a specific design principle in the field of statistics education, for example, the idea of *growing samples* (Bakker, 2004; Ben-Zvi, Aridor, Makar, & Bakker, 2012; Konold & Pollatsek, 2002), or approaches related to computer-based learning, inquiry-based learning, problem-based teaching or Realistic Mathematics Education (RME).[3] According to diSessa and Cobb (2004), some of these approaches can be labelled *frameworks for action*, whereas others can be labelled *domain-specific instructional theories*. They typically have a prescriptive nature and address the design of learning processes and learning environments (see Chap. 16). The question whether they can be considered *foreground or background theories* (Mason & Waywood, 1996) strongly depends on the purpose of the investigations.

In this category, we also note theories used to understand *language* and its influence in statistics learning, *contexts* and their influence on students' understanding and their learning process, *technology use* in the mathematics learning process and—in a broader sense—*representations* and their role in the learning of statistics.

[3] We refer the reader to Chap. 16 of this volume for a discussion of RME.

Furthermore, topics such as how to teach in relation to gender (Yolcu, 2014) or blind children (Vita & Kataoka, 2014) were grounded on this kind of theory.

11.4.1.5 Theories from Mathematics or Science Education (TMSE)

This category contains theories in mathematics (or science) education that are being used in SER. These theories are mainly used as background theories (Mason & Waywood, 1996) because they constitute a background for which objects or concepts are being studied in the empirical investigations. Generally, they have a descriptive or analytic nature, but, often, they also imply general advice for teaching.

Theories in mathematics education are about epistemological questions concerning, for example, learning and understanding mathematics. For instance, Cobb and Bauersfeld's (1995) and Cobb, Yackel, and Wood's (1992) "translation" of interactionism (Blumer, 1986) into mathematical education constitutes an example of a theorization belonging to this category, as they conceptualize interaction from a mathematics education perspective.

Basic topics that these theories deal with are ideas about contexts and context use in mathematics education, ideas such as *(guided) reinvention* (Freudenthal, 1973, 1991), *Realistic Mathematics Education* (Gravemeijer & Doorman, 1999) or Cobb, Yackel, and Wood's (1989) *socio-mathematical norms*. What is also referred to are theories about *language in science* (Lemke 1990) or Vygotsky's (2001) idea about *everyday and scientific concepts* in the context of mathematics and science education.

11.4.1.6 Theories with a Broader Range on Epistemological Aspects (TEA)

Theories in this category concern learning or cognitive development from a perspective that is not restricted to mathematics or statistics education and that has its origin in another discipline, such as psychology, sociology or philosophy. To these belong, for instance, Vygotsky's (1978) *learning theory*, Bourdieu's (1984) understanding of *culture*, von Glasersfeld's (1995) *constructivism* or Biggs and Collis' (1982) *SOLO model* as a cognitive developmental model.

These theories constitute a broader theoretical background for the investigations. Therefore they are used as background theories (Mason & Waywood, 1996) and refer to epistemological basic ideas that are fundamental for the orientation of the investigations. According to diSessa and Cobb's (2004) distinction, they can be considered *grand theories*, which provide basic assumptions on, for example, learning or *orienting frameworks*, which help to specify issues of learning and implications for teaching.

Table 11.1 Frequencies and percentages of articles referring to each theoretical approach

	# of articles	SPdT	SPcT	TDF	TMSE	TEA
ESM	9	9	4	4	4	2
JRME	2	2	0	0	0	0
MTL	6	3	5	5	5	2
SERJ	18	12	9	11	7	6
Total	35 (100%)	26 (74%)	18 (51%)	20 (57%)	16 (46%)	10 (29%)

Table 11.2 Frequencies and percentages of how articles use different theoretical approaches

	# of articles	Group 1 1 type of theory	Group 2 2 types of theory	Group 3 3 types of theory	Group 4 4 types of theory	Group 5 5 types of theory
ESM	9	1	4	1	3	0
JRME	2	2	0	0	0	0
MTL	6	0	1	2	3	0
SERJ	18	1	10	4	3	0
Total	35 (100%)	4 (11%)	15 (43%)	7 (20%)	9 (26%)	0 (0%)

11.4.2 Types and Combinations of Theories Used for Grounding SER

In the following, we outline to what extent the articles in our review referred to different types or combinations of theories. When answering our second research question, we differentiated between the five categories of theories characterized in the previous section.

Table 11.1 shows how many articles of each journal were integrated in our review (e.g., nine articles from ESM) and shows per row how many of these articles referred to theories of the above-mentioned categories.

Statistical Product Theories were most often used (74%) in the articles considered in our review. Examples of these theories are Konold and Pollatsek's (2002) conceptual understanding of signal and noise (used in, e.g., Bakker, 2004), Watson & Callingham's (2003) hierarchical developmental model of understanding of variation (used, e.g., in Watson, Callingham, & Kelly, 2007), Jones, Langrall, Thornton, & Mogill's, (1997) four-stage cognitive developmental model on elementary school students' probabilistic thinking (based on levels of thinking) or Polaki's (2002) cognitive developmental model on elementary school students' learning of probability.

Theories with a Broader Range on Epistemological Aspects (TEA) are apparent in 29% of the articles. Hence, in 71% of the articles, the studies are not referring to or being underpinned by general learning theories.

Related to the topic of *networking of theories* (Bikner-Ahsbahs & Prediger, 2014), another perspective in our review was on the question in how far the articles used different kinds of theories. The results of this analysis are displayed in Table 11.2.

Table 11.2 shows that only 4 out of 35 articles draw on one single theory. Keeping to the topic of statistics, these articles concerned *Statistical Product Theories (SPdT)* or *Statistical Process Theories (SPcT)*. For instance, in Rubel's (2007) article about "Middle school and high school students' probabilistic reasoning on coin tasks", all theoretical approaches were marked as Statistical Product Theories, as they dealt with students' conception of key concepts of probability such as sample space, independent event and probability comparison.

The largest group of articles (43%) used two different types of theories. Within this group, the most common combination (4 out of 14) was a *Statistical Process Theory (SPcT)* with a *didactical theory (TDF)*, such as teaching issues in relation to gender (Yolcu, 2014), blind students (Vita & Kataoka, 2014) and technological tools (Watson, 2008; Paparistodemou & Meletiou-Mavrotheris, 2008). However, in these cases, the didactical dimension was not explicitly grounded in a broader, general framework. Statistical Process Theories/frameworks constitute the core of the theoretical framing. What also occurred was a combination of a *Statistical Product Theory (SPdT)* and a *Statistical Process Theory (SPcT)* as well as a combination of a *Statistical Product Theory (SPdT)* and a *Theory with a Broader Range on Epistemological Aspects (TEA)* (each 3 out of 14). In the latter case, Statistical Product Theory was framed by a more general theoretical framework. For example, in Reading (2004), Biggs and Collis' (1982) SOLO model was used in order to frame Mooney's (2002) Statistical Thinking Framework, which is based on the SOLO model. Other combinations of different kinds of theories did not occur more often than two times.

Whereas 20% of the articles referred to three different kinds of theories, in 26% of the cases, four different types of theories were used. These included the types *Theories in Mathematics and Science Education (TMSE)* and *Theories with a Broader Range on Epistemological Aspects (TEA)*. Finally, there were no articles in this review that combined theories of all five types.

11.5 Recommendations for the Future of Theory Development in SER

Our review revealed that SER has begun to mature as a scientific discipline in terms of using domain-specific theories in grounding and guiding research. Focusing on domain-specific theories of statistics is of course important and necessary; it is what gives SER its identity. However, we suggest that SER could benefit from extending and strengthening its use of background theories and orienting frameworks. The book edited by Koschmann (2011) shows how scholars from different disciplines theoretically reflected on an example at the boundary of science and statistics education. Many crucial epistemological issues and phenomena in SER are often treated implicitly or neglected. We therefore encourage statistics educators and researchers to be explicit about their background theories and orienting frameworks.

We continue with a list of epistemological issues that we consider in need of further theoretical elaboration. The list entails topics of study that we think need a stronger theoretical basis or treatment. We summarize our recommendations as follows and elaborate on them in the remainder of this section:

1. More explicit attention needs to be paid to how students can learn historically developed disciplinary, formal knowledge. More explicit attention on theorizing the relationship between formal and personal views of statistics will help to move the field forward (beyond statistics education) (Makar, 2014).
2. In addition to static categorizations of student thinking, we need insights into the dynamics between categories or levels. In the case of valid types of reasoning: How can combinations or even integration be promoted? In the case of levels: How can transitions to higher levels be promoted?
3. There is a need for more fundamental theories on the impact of digital technology on learning statistics but also on how to teach with digital technology. Reflection on how the nature of statistical knowledge itself changes due to such technology will also be necessary (see, e.g., Biehler, Ben-Zvi, Bakker, & Makar, 2013; Gould, 2010).
4. There is a need for a deeper theoretical conceptualization of context and contextualizing in statistics education (see, e.g., Bakker & Derry, 2011; Gil & Ben-Zvi, 2011; Makar & Ben-Zvi, 2011).
5. Consider potential benefits of a semantic theory that has been proposed as underpinning research on statistical inference: inferentialism. We do not want to suggest this is the only or best way forward, but it is in our view an interesting candidate to shed a new light on long-standing issues (e.g., Bakker, Ben-Zvi, & Makar, 2017).

11.5.1 The Relationship Between Formal and Personal Views of Statistics

We examined how statistical knowledge has been approached from the disciplinary perspective of statistics and/or from the empirical perspective of students. In developing statistics education, both approaches are important and reflexively dependent. On the one hand, the statistical perspective defines the nature of the discipline and provides guidance for formulating standards and teaching goals. On the other hand, empirical research of students' understandings and dealings with statistics becomes essential in a pedagogy that emphasizes that teaching builds students' learning trajectories or progressions to the goals of the standards from where the students are (see also Chap. 9, this volume). The idea of building teaching from where the students are, and attempting to align students' conceptions with the target of teaching, is implicit in many of the studies of our mapping that focused on characterizing students' conceptions or prior knowledge (e.g., Dierdorp, Bakker, Eijkelhof, & van Maanen, 2011; Lee, Angotti, & Tarr, 2010; Watson, 2009).

Designing instruction that builds on students' understandings (Jones, Langrall, & Mooney, 2007; Konold, Harradine, & Kazak, 2007) and aligning students' conceptions with the target of teaching come with a "pedagogic challenge" (Pratt, 2005, p. 175) (see also Chap. 16 in this volume). Much evidence testifies that students' prior understanding often impacts (Sharma, 2014) and stands in conflict with the formal way of understanding key concepts of probability and statistics (Fischbein & Schnarch, 1997; Kahneman, Slovic, & Tversky, 1982). What is the meaning, or possibility, to build on personal ideas when they are at odds with the learning goals? How is it possible to align opposing understandings? It is in relation to such questions we claim that the pedagogic challenge Pratt (2005) points to should be considered. It is in relation to such question that the relationship between formal and personal meanings of statistics has not yet had the theoretical treatment it needs.

In a similar vein, if we are to develop frameworks for action (diSessa & Cobb, 2004) and hypothetical learning trajectories (Simon, 1995), which are based on connecting formal and personal meanings, there is a need of a theoretical treatment that makes it possible to understand, explain and predict processes involved in this relationship. Where some authors emphasize the continuity of such development (Abrahamson, 2012), others consider its discontinuity (Yerushalmy & Chazan, 2008). We speculate that constructivists prefer to stress continuity (because students have to construct new knowledge on the basis of old knowledge) and that sociocultural theorists can more easily live with discontinuities (because there are multiple distinct practices in which students learn to participate). What these views seem to share are explanations in one direction: from student to discipline (students constructing disciplinary knowledge) or from discipline to student (internalizing sociocultural practices). An alternative approach, expressed by Rosen, Palatnik and Abrahamson (2016), is to work from a middle ground in both directions. In this respect, there seems to be a renewed need to reconcile the acquisition and participation metaphors of learning (Sfard, 1998; Taylor, Noorloos, & Bakker, in press).

11.5.2 Static Versus Dynamic Aspects of Frameworks About Processes

The cognitive frameworks used in research on students' understanding of statistical products often specify levels in students' understandings. They are at times descriptive in nature as they provide an overview of different ways of understanding key concepts in statistics and probability. Cognitive frameworks have commonly been used for prescriptive purposes: They have proven useful for teachers to design, implement and assess learning environments in statistics and probability (Jones et al., 2007). Of course, we find such results promising. However, according to the discussion in the preceding paragraph, the articles involved in our mapping provide little explanatory power of understanding processes involved in moving from one level in the framework to another level, neither do they give explanations for why students may respond according to one level in one situation and on another level in another situation. From a design-research perspective (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003), a cognitive framework may provide input for formulating a hypothetical learning trajectory (cf., Bakker, 2004). But such frameworks provide limited guidance for the means to support such a learning trajectory and how the levels of the students' responses depend on affordances provided by the learning environment such as interaction with teachers, other students and mediating tools (Ryve, Larsson, & Nilsson, 2013; Sfard, 2008). Hence, a way to strengthen the power of a cognitive framework could be to underpin it by an explicit background theory that orients researchers to conceptualize transitions between levels and the means of supporting such transitions.

Apart from Watson and Callingham's (2003) framework, we do not know of frameworks of statistical processes that really highlight issues of learning in terms of progression within or between the different categories of the framework. Frameworks are typically described in general terms according to practice developed in the statistical discipline without attention to how categories can be combined or higher levels can be reached. Defining clear boundaries around categories has been argued to be an important endeavour in scientific research (Bowker & Star, 2000), but in educational research, we also need insight into how to cross boundaries.

For instance, several of the articles in our review (e.g., Ben-Zvi, 2004; Lee et al., 2010; Watson et al., 2007; Watson & Kelly, 2004) refer to Wild and Pfannkuch's (1999) framework of statistical thinking (SPcT). Several key components of statistical investigations are highlighted in this framework, for example, (1) the ability to formulate a statistical researchable question, (2) modelling and (3) contextual awareness. However, in the articles of the review, components like these three were dealt with only in general (or implicit) terms. Readers were provided with little information of how to understand qualitative differences in how students expressed aspects of the components.

Say that a group of students are interested in whether there is a "real" difference in expected values between girls and boys in scoring goals in penalty kicks in soccer. The class comes up with a number of different formulations of the exact statistical question to investigate. The question we ask is how can research provide tools for the teacher to conceptualize and distinguish progression of statistical sophistication in the students' questions? We also ask by which means the teacher should act in order to develop students' ability to formulate statistical research questions of a certain quality. Say that one student limits the research question to examine only right-footed players while another student includes both right-footed and left-footed players. Related to data collection, one student may say that they need to take into consideration if the penalty is made during practice or during match. As we can see, the situation quickly becomes very complex. We wonder whether it is possible to develop principles that teachers can use as a guide to conceptualize qualities and progression within and between the different components of a statistical investigations and, subsequently, principles for how progression of students' understanding of statistical processes can be conceptualized and supported.

To deal with such learning issues, we think that the frameworks of statistical processes need to be further grounded in theories of learning. In the same spirit, it would be interesting to see how learning, according to a framework describing statistical processes, such as Makar and Rubin's (2009) framework of informal inferential reasoning, can be conceptualized in different ways by adapting the framework to different background theories of learning, such as constructivism, sociocultural perspective, symbolic interactionism and distributed cognition (cf., Cobb, 2007).

11.5.3 Theories About Technology Use

Within the category of didactical theories (TDF), there was an emphasis on technology (e.g., Prodromou & Pratt, 2006; dos Santos Ferreira, Yumi Kataoka, & Karreer, 2014; Lee et al., 2010). However, in the reviewed articles, there was no deeper theorization of computer-assisted instruction in statistics. We did not find what diSessa and Cobb (2004) would describe as frameworks for action or domain-specific learning theories. The research motivates and conceptualizes new technology mainly on empirical results, emerging from individual case studies. It is hard to note any accumulated results or consensus, except for an overall argumentation of the possibility to provide visualizations, simulations and different forms of representations by new technology. In the 35 articles, we found no theoretical attempts on a more specific level with the intention to provide prescriptive information for supporting statistics learning with technology, such as guiding principles for designing tasks and sequencing tasks in a digital learning environment or frameworks for explaining and understanding the relationship between digital and analogue learning environments. Watson (2008) comes close in her discussion of boxplots in TinkerPlots. Of course, outside the review sample, there are more studies on this topic (e.g., Ben-Zvi, 2000; Biehler et al., 2013; Konold & Kazak, 2008).

Theories from mathematics education such as about instrumental genesis (Drijvers & Trouche, 2008) could also be useful to think more fundamentally how technology may affect learning. But there is more: Technology changes the discipline of statistics itself and to the need to rethink learning goals of statistics education (Gould, 2010).

11.5.4 Context and Contextualization

Like most disciplines, statistics is traditionally seen as theoretical, something general applied in practical contexts. In line with this idea, Wild and Pfannkuch (1999, p. 28) talked about the importance of "shuttling between the contextual and statistical spheres." A special issue in MTL in 2011 (Makar & Ben-Zvi, 2011) indicated that the SER community has considered this topic of the relation between statistics and context significant.

Articles that addressed issues of context were not restricted to just one category in the review. Context was addressed according to TEA (e.g., Halldén's (1999) theory of contextualization), TMSE (e.g, context connected to the idea of guided reinvention (Dierdorp et al., 2011)) and TDF (e.g., context as a means of support (Pfannkuch, 2011)).

The attention to context and contextualizing in SER reflects an increased awareness of the situated and social nature of learning and teaching in general (Kirshner & Whitson, 1997) and in mathematics in particular (Lerman, 2000). However, taking a closer look into our review, we see that situated, interactive and contextual issues of learning and teaching of statistics are often dealt with implicit and informal ways. Of the 35 articles, only 8 explicitly theorized the meaning and role of context and contextualization in the learning of statistics; of these eight, four were in the special issue "the Role of Context in Developing Reasoning about Informal Statistical Inference" in MTL 2011 (Makar & Ben-Zvi, 2011). So, if we assume that contextual issues are essential to take into account in understanding learning and teaching of statistics, our review points to the need for stronger theoretical grounding. An increased theoretical basis and clarity will not only enable researchers to build on one another's work in a reliable way, but it will also increase the implementation fidelity of research outcomes into a classroom practice (Lester, 2010; Silver & Herbst, 2007).

11.5.5 Inferentialism in Statistics Education

A last point concerns a theory that increasingly is attracting attention in statistics education (and many other disciplines): *inferentialism*. Inferentialism is a semantic theory, formulated by a philosopher (Brandom, 2000), which puts inference at the core of human knowing and thus fits well with the idea of statistical inference at the heart of statistical knowing. More generally, Brandom privileges inference over representation. This means that in his view, the ability to represent anything rests on practices of human reasoning. The opposite route of explanation has been more common in the history of philosophy and education: Once we can represent we can reason. This is a representationalist view, which has been criticized by many philosophers (Rorty, 1979) and educational researchers (Cobb, Yackel, & Wood, 1992). Bakker and Derry (2011) argued that there is also a risk in statistics education to adopt, without being aware of it, a representationalist view, which is to assume that once students know the key representations of statistics, they can reason statistically. What we often see in statistics curricula as a consequence is an atomistic approach: mean, median, mode, range and standard deviation are dealt with one by one. And the idea of distribution is only introduced once concepts and representations such as mean, standard deviation and Cartesian graphs are covered.

Aiming to counterbalance such pitfalls and tendencies, Bakker and Derry (2011) draw three lessons from inferentialism for statistics education. First, statistical concepts should in their view be primarily understood in inferential terms, that is, in their role in reasoning. Trying to move carefully from descriptive to prescriptive ideas, they explicitly take the step from philosophy to education: If, from a philosophical point of view, the inferential role of concepts should be privileged over their representational function, then educators may also need to emphasize the importance of concepts in use. This is the first lesson drawn from inferentialism. The second lesson is that a holistic approach should be prioritized over an atomistic one. Given that concepts only have meaning in relation to other concepts, statistical concepts should be learned in packages—in relation to each other. For example, mean and standard deviations have more meaning in relation to distribution than in isolation. This implies that informal attention to distribution may be needed well before any formal definition can be given (Bakker, 2004). As a third lesson, Bakker and Derry (2011) illustrate what privileging an inferentialist approach to teaching statistics may look like in contrast to a representationalist approach. In this way, they try to link a theoretical background theory on epistemology to didactical ideas about informal inferential reasoning. As their study testified, such theoretical work is far from trivial but, in our view, necessary.

In our view, inferentialism has the potential to address the previous needs from a fresh perspective. Firstly, it offers a perspicuous view on the *relation between the individual and social* (Schacht & Hußmann, 2015) that underlies the pedagogic challenge formulated in the first need. Secondly, by understanding concepts, categories and representations in terms of inference and reasoning, the inferentialist language and ways of thinking may well offer the *dynamic and holistic view* that can help to avoid static usage of frameworks with categories or levels. Thirdly, the issue of technology forces scholars to think about the distributed cognition (Hutchins, 1995) among humans and machines. Although Brandom's primary interest is *human* reasoning, the focus on inferences can still offer a fresh perspective on what students need to learn. When using technology, particular inferences are outsourced in computational form to technology, but humans still have to decide which technology to use and how to interpret the outcomes (Hoyles, Noss, Kent, & Bakker, 2010). This can be challenging because of the black box nature of much digital technology. In the travelling metaphor used by Biehler et al. (2013), doing statistics by hand is like walking—step by step with attention to many details. Doing statistics with technology allows us to travel fast and far, with the obvious advantage of being able to infer things that are impossible with pen and article but with the drawback of not having sight on the route taken.

Lastly, inferentialism has also been used to zoom in and conceptualize contextual issues in learning mathematics and other disciplinary knowledge. Heusdens, Bakker, Baartman and De Bruijn (2015) have proposed to use the term *contextualizing* both for bringing ideas and actions into a theoretical context (creating conceptual coherence by means of conceptualizing) and for bringing them into practical context (concretizing as relating general ideas to specific situations or actions). For statistics, this implies that one should not see statistics as decontextualized knowledge but as a discipline that brings statistical ideas and techniques into a conceptual context but also relates them to concrete situations. As Heusdens et al. (2015) illustrate with culinary examples, these processes of conceptualizing and concretizing can occur simultaneously, which suggests a way out of the aforementioned dichotomy. They use inferentialism, in particular the idea of *webs of reasons*, to highlight the similarity between conceptualizing and concretizing: In any concrete situation, multiple reasons are at stake. Some may be theoretical and some practical. Some may be statistical and others pragmatic. Although there is still a long way to better understand the complexity of the interplay between various types of reasons, we think that inferentialism offers a helpful theoretical lens that can illuminate how students can learn to contextualize and integrate statistical and contextual considerations.

11.6 Conclusions

The goal of the review presented here was to categorize types of theories used in Statistics Education Research (SER). The review addressed the following research questions:

1. What is the nature of theories used in Statistics Education Research?
2. To what extent are different types and combinations of theories used in Statistics Education Research?

In response to our first research question, we distinguished four different main categories: *Theories of Statistics in Statistics Education Research*, where some could be characterized as being focused on products (SPdT) and others as process oriented (SPcT). Some theories focused on statistics itself and others on student learning of particular statistical content. *Theories with a Didactical Focus (TDF)* were predominantly foreground theories on instructional activities, technology or language. *Theories in Mathematics or Science Education (TMSE)* contained background theories, for example, on Realistic Mathematics Education, socio-mathematical norms or everyday versus scientific concepts. And finally, *Theories with a Broader Range on Epistemological Aspects (TEA)* stemmed from, for example, Vygotsky (1978), Bourdieu (1984), von Glasersfeld (1995) and Biggs and Collis (SOLO) (1982).

In response to our second research question, we concluded that many authors of the articles in our study used several types of theories. We do not mean to demean careful usage and development of a single theory or to imply that including more theories is always a better approach. Yet we find drawing on multiple theories promising because, in our experience, different theoretical resources are typically needed to study complex issues in depth (cf. diSessa & Cobb, 2004).

Last, we moved from a descriptive review perspective to a more critical stance. We recommended five themes that in our view need further thought and theoretical treatment by statistics education researchers.

As indicated, this review study has its limitations. We hope others feel invited by this modest review of how theories have been used in recent empirical SER articles to deal with the topic more extensively. For example, a larger set of publications, including theoretical ones, may help to identify particular trends. Furthermore, reading of the full articles will help to judge how well theories are put to work. Within mathematics education, theory has been the topic of many publications including books (Bikner-Ahsbahs & Prediger, 2014; Sriraman & English, 2010). However, in SER such publications are practically absent. Promising approaches may be to use one study in statistics education as the source of reflection from different theoretical perspectives (cf. Bikner-Ahsbahs & Prediger, 2014; Koschmann, 2011).

References

Abrahamson, D. (2012). Seeing chance: Perceptual reasoning as an epistemic resource for grounding compound event spaces. In R. Biehler & D. Pratt (Eds.), Probability in reasoning about data and risk (special issue). *ZDM, 44*(7), 869–881.

Assude, T., Boero, P., Herbst, P., Lerman, S., & Radford, L. (2008). *The notion and roles of theory in mathematics education research*. Paper presented at The 10th International Congress on Mathematics Education, Monterrey, Mexico, July 6-13.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, the Netherlands: CD Beta Press.

Bakker, A. (2014). Implications of technology on what students need to know about statistics. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen - using tools for learning mathematics and statistics* (pp. 143–152). Wiesbaden, Germany: Springer.

Bakker, A., & Akkerman, S. F. (2014). A boundary-crossing approach to support students' integration of statistical and work-related knowledge. *Educational Studies in Mathematics, 86*(2), 223–237.

Bakker, A., Ben-Zvi, D., & Makar, K. (2017). An inferentialist perspective on the coordination of actions and reasons involved in making a statistical inference. *Mathematics Education Research Journal*, 455–470. https://doi.org/10.1007/s13394-016-0187-x

Bakker, A., & Gravemeijer, K. P. (2006). An historical phenomenology of mean and median. *Educational Studies in Mathematics, 62*(2), 149–168.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1-2), 5–26.

Bakker, A., & Smit, J. (2017). Theory development in design-based research: An example about scaffolding mathematical language. In S. Doff & R. Komoss (Eds.), *How does change happen? Wandel im Fachunterricht analysieren und gestalten* (pp. 109–124). Wiesbaden, Germany: Springer.

Bakker, A., & van Eerde, H. A. A. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Doing qualitative research: Methodology and methods in mathematics education* (pp. 429–466). Berlin, Germany: Springer.

Bansilal, S. (2014). Using an APOS framework to understand teachers' responses to questions on the normal distribution. *Statistics Education Research Journal, 13*(2), 42–57.

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1–2), 127–155.

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal, 3*(2), 42–63.

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM, 44*(7), 913–925.

Ben-Zvi, D., Gil, E., & Apel, N. (2007). What is hidden beyond the data? Helping young students to reason and argue about some wider universe. In D. Pratt & J. Ainley (Eds.), *Reasoning about Informal Inferential Statistical Reasoning: A collection of current research studies. Proceedings of the 5th International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL5)* (pp. 29–35). Warwick, UK: University of Warwick.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In A. Bishop, K. Clement, C. Keitel, J. Kilpatrick, & A. Y. L. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–689). New York: Springer.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The solo taxonomy*. New York: Academic Press.

Bikner-Ahsbahs, A., & Prediger, S. (2010). Networking of theories - an approach for exploiting the diversity. In B. Sririman & L. English (Eds.), *Theories of mathematics education: Seeking new frontiers* (pp. 483–512). Berlin: Springer.

Bikner-Ahsbahs, A., & Prediger, S. (Eds.). (2014). *Networking of theories as a research practice in mathematics education*. Cham: Springer.

Blumer, H. (1986). *Symbolic interactionism: Perspective and method*. Berkeley: University of California Press.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Cambridge, MA: Harvard University Press.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT press.

Brandom, R. (2000). *Articulating reasons: An introduction to inferentialism*. Cambridge, MA: Harvard University Press.

Brown, A. L., & Campione, J. C. (1994). *Guided discovery in a community of learners*. Cambridge, MA: MIT Press.

Cobb, P. (2007). Putting philosophy to work. Coping with multiple theoretical perspectives. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 3–38). Greenwich, CT: Information Age Publishing.

Cobb, P., & Bauersfeld, H. (1995). *The emergence of mathematical meaning: Interaction in classroom cultures*. Hillsdale, NJ: Erlbaum.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

Cobb, P., Yackel, E., & Wood, T. (1989). Young children's emotional acts while doing mathematical problem solving. In D. McLeod & V. M. Adams (Eds.), *Affect and mathematical problem solving: A new perspective* (pp. 117–148). New York, NY: Springer.

Cobb, P., Yackel, E., & Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education, 23*(1), 2–33.

Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*(1–2), 132–151.

Dierdorp, A., Bakker, A., van Maanen, J. A., & Eijkelhof, H. M. C. (2014). Meaningful statistics in professional practices as a bridge between mathematics and science: An evaluation of a design research project. *International Journal of STEM Education, 1*(1), 1–15.

diSessa, A. A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *The Journal of the Learning Sciences, 13*(1), 77–103.

diSessa, A. A., & Sherin, B. L. (2000). Meta-representation: An introduction. *The Journal of Mathematical Behavior, 19*(4), 385–398.

dos Santos Ferreira, R., Yumi Kataoka, V., & Karreer, M. (2014). Teaching probability with the support of the R statistical software. *Statistics Education Research Journal, 13*(2), 132–147.
Drijvers, P., & Trouche, L. (2008). From artifacts to instruments: A theoretical framework behind the orchestra metaphor. In G. W. Blume & M. K. Heid (Eds.), *Research on technology and the teaching and learning of mathematics*, *Cases and perspectives* (Vol. 2, pp. 363–392). Charlotte, NC: Information Age.
Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher, 31*(8), 4–14.
Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*(1), 96–105.
Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, The Netherlands: Reidel.
Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht, The Netherlands: Reidel.
Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Gal, I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 39–64). New York: Springer.
Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1–2), 87–108.
Gough, D., Oliver, S., & Thomas, J. (2013). *Learning from research: Systematic reviews for informing policy decisions*. London: University of London EPPI-Centre.
Gould, R. (2010). Statistics and the modern student. *International Statistical Review, 78*(2), 297–315.
Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics, 39*(1–3), 111–129.
Griffith, J. D., Adams, L. T., Gu, L. L., Hart, C. L., & Nichols-Whitehead, P. (2012). Students' attitudes toward statistics across the disciplines: A mixed-methods approach. *Statistics Education Research Journal, 11*(2), 45–56.
Groth, R. E. (2015). Working at the boundaries of mathematics education and statistics education communities of practice. *Journal for Research in Mathematics Education, 46*(1), 4–16.
Halldén, O. (1999). Conceptual change and contextualization. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 53–65). Oxford: Pergamon, Elsevier Science.
Heusdens, W. T., Bakker, A., Baartman, L. K. J., & De Bruijn, E. (2015). Contextualising vocational knowledge: A theoretical framework and illustrations from culinary education. *Vocations and Learning, 9*(2), 151–165.
Hoyles, C., Noss, R., Kent, P., & Bakker, A. (2010). *Improving mathematics at work: The need for techno-mathematical literacies*. Abingdon: Routledge.
Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
Jones, G., Langrall, C., Thornton, C., & Mogill, T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32*(2), 101–125.
Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–956). Greenwich: NCTM.
Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
Kirshner, D., & Whitson, J. A. (Eds.). (1997). *Situated cognition theory: Social, semiotic, and neurological perspectives*. Hillsdale, NJ: Erlbaum.
Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning, 12*(3), 217–230.

Konold, C., Higgins, T., Russell, S., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1). http://escholarship.org/uc/item/38p7c94v.

Koschmann, T. D. (2011). *Theories of learning and studies of instructional practice*. New York, NY: Springer.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lee, H. S., Angotti, R. L., & Tarr, J. E. (2010). Making comparisons between observed data and expected outcomes: students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal, 9*(1), 68–96.

Lemke, J. (1990). *Talking science: Language, learning, and values.* Norwood, N.J.: Ablex.

Lerman, S. (2000). The social turn in mathematics education research. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 19–44). Westport, CN: Ablex.

Lerman, S., Xu, G., & Tsatsaroni, A. (2002). Developing theories of mathematics education research: The ESM story. *Educational Studies in Mathematics, 51*(1–2), 23–40.

Lester, F. K. (2010). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. In B. Sriraman & L. English (Eds.), *Theories of mathematics education: Seeking new frontiers* (pp. 67–85). Heidelberg: Springer.

Lewin, K. (1951). In D. Cartwright (Ed.), *Field theory in social sciences; Selected theoretical papers*. New York: Harper and Row.

Makar, K. (2014). Young children' s explorations of average through informal inferential reasoning. *Educational Studies in Mathematics, 86*(1), 61–78.

Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 1–4.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.

Marton, F., Runesson, U., & Tsui, A. (2004). The space of learning. In F. Marton & A. Tsui (Eds.), *Classroom discourse and the space of learning* (pp. 3–40). Mahwah: Erlbaum.

Mary, C., & Gattuso, L. (2005). Trois problèmes semblables de moyenne pas si semblables que ça! L'influence de la structure d'un problème sur les réponses des élèves. *Statistics Education Research Journal, 4*(2), 82–102.

Mason, J., & Waywood, A. (1996). The role of theory in mathematics education and research. In A. J. Bishop, M. A. K. Clements, C. Keitel-Kreidt, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 1055–1089). Kluwer: Dordrecht.

McKenney, S., & Reeves, T. C. (2012). *Conducting educational design research*. London: Routledge.

Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning, 4*(1), 23–63.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics: Reston. VA.

Nilsson, P. (2009). Conceptual variation and coordination in probability reasoning. *The Journal of Mathematical Behavior, 28*(4), 247–261.

Niss, M. (2007). Reflections on the state of and trends in research in mathematics teaching and learning. From here to utopia. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1293–1312). Greenwich: NCTM.

Paparistodemou, E., & Meletiou-Mavrotheris, M. (2008). Developing young students' informal inference skills in data analysis. *Statistics Education Research Journal, 7*(2), 83–106.

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning, 13*(1–2), 27–46.

Pfannkuch, M., & Wild, C. J. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 17–46). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Polaki, M. (2002). Using instruction to identify key features of Basotho elementary students' growth in probabilistic thinking. *Mathematical Thinking and Learning, 4*(4), 285–313.

Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education, 31*(5), 602–625.

Pratt, D. (2005). How do teachers foster students' understanding of probability? In G. Jones (Ed.), *Exploring probability in school* (pp. 171–189). New York: Springer.

Prodromou, T., & Pratt, D. (2006). The role of causality in the co-ordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal, 5*(2), 69–88.

Reaburn, R. (2014). Introductory statistics course tertiary students' understanding of p-values. *Statistics Education Research Journal, 13*(1), 53–65.

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal, 3*(2), 84–105.

Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton: Princeton University Press.

Rosen, D. M., Palatnik, A., & Abrahamson, D. (2016). Tradeoffs of situatedness: Iconicity constrains the development of content-oriented sensorimotor schemes. In M. Wood, E. Turner, & M. Civil (Eds.), *Sin fronteras: Questioning borders with(in) mathematics education. Proceedings of the 38th Annual Meeting of the North-American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA)*, *Technology* (Vol. 12, pp. 1509–1516). Tucson, AZ: University of Arizona.

Rubel, L. H. (2007). Middle school and high school students' probabilistic reasoning on coin tasks. *Journal for Research in Mathematics Education, 38*(5), 531–556.

Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics [CD-ROM]*. Voorburg: The Netherlands: International Statistical Institute.

Ryve, A., Larsson, M., & Nilsson, P. (2013). Analyzing content and participation in classroom discourse: Dimensions of variation, mediating tools, and conceptual accountability. *Scandinavian Journal of Educational Research, 57*(1), 101–114.

Säljö, R. (2003). Epilogue: From transfer to boundary-crossing. In T. Tuomi-Gröhn & Y. Engeström (Eds.), *Between school and work. New perspectives on transfer and boundary-crossing* (pp. 311–321). Amsterdam: Pergamon.

Säljö, R. (2011). On plants and textual representations of plants. Learning to reason in institutional categories. In T. Koschmann (Ed.), *Theories of learning and studies of instructional practice* (pp. 279–290). New York: Springer.

Schacht, F., & Hußmann, S. (2015 July). Between the social and the individual: Reconfiguring a familiar relation. *Philosophy of Mathematics Education Journal, 29*, 1–26.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher, 27*(2), 4–13.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge, UK: Cambridge University Press.

Sharma, S. (2014). Influence of culture on secondary school students' understanding of statistics: A Fijian perspective. *Statistics Education Research Journal, 13*(2), 104–117.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Charlotte, NC: Information Age Publishing.

Silver, E., & Herbst, P. (2007). Theory in mathematics education scholarship. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 39–67). Greenwich: NCTM.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114–145.

Sriraman, B., & English, L. (2010). Surveying theories and philosophies of mathematics education. In B. Sriraman & L. English (Eds.), *Theories of mathematics education: Seeking new frontiers* (pp. 7–32). Berlin: Springer.

Steinbring, H. (2008). Changed views on mathematical knowledge in the course of didactical theory development. *ZDM, 40*(2), 303–316.

Straesser, R. (2014). Stoffdidaktik in mathematics education. In S. Lerman (Ed.), *Encyclopaedia of mathematics education* (pp. 566–570). Dordrecht: Springer.

Swedish National Agency for Education. (2012). *Curriculum for the compulsory school system, the pre-school class and the leisure-time centre 2011*. Stockholm: Swedish National Agency for Education.

Taylor, S. D., Noorloos, R., & Bakker, A. (in press). Mastering as an inferentialist alternative to the acquisition and participation metaphors of learning. *Journal of Philosophy of Education*.

Vita, A. C., & Kataoka, V. Y. (2014). Blind students' learning of probability through the use of a tactile model. *Statistics Education Research Journal, 13*(2), 148–163.

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. London: Falmer.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (2001). *A Construção do Pensamento e da Linguagem [The Construction of Thought and Language]. Translated by P. Bezerra*. São Paulo: Martins Fontes.

Watson, J. M. (1997). Assessing statistical thinking using the media. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). Amsterdam, The Netherlands: IOS Press.

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3–46.

Watson, J., & Callingham, R. (2014). Two-way tables: Issues at the heart of statistics and probability for students and teachers. *Mathematical Thinking and Learning, 16*(4), 254–284.

Watson, J. M. (2008). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal, 7*(2), 59–82.

Watson, J. M. (2009). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal, 8*(1), 32–61.

Watson, J. M., Callingham, R. A., & Kelly, B. A. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical Thinking and Learning, 9*(2), 83–130.

Watson, J. M., & Kelly, B. A. (2004). Statistical variation in a chance setting: A two-year study. *Educational Studies in Mathematics, 57*(1), 121–144.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistics Review, 67*(3), 223–265.

Xu, W., Zhang, Y., Su, C., Cui, Z., & Qi, X. (2014). Roles of technology in student learning of university level biostatistics. *Statistics Education Research Journal, 13*(1), 66–76.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458–477.

Yerushalmy, M., & Chazan, D. (2008). Technology and curriculum design: The ordering of discontinuities in school algebra. In L. English et al. (Eds.), *Handbook of international research in mathematics education* (2nd ed., pp. 806–837). London: Routledge.

Yolcu, A. (2014). Middle school students' statistical literacy: Role of grade level and gender. *Statistics Education Research Journal, 13*(2), 118–131.

Chapter 12
Reimagining Curriculum Approaches

Maxine Pfannkuch

Abstract As new societal learning goals are formulated and people and technology shape, grow, and challenge statistical practice and thinking, educators respond through researching, imagining, and implementing new curriculum approaches. In our reimagining of curriculum approaches, we have chosen to discuss learning experiences that all students could engage in as part of their enculturation into thinking from a statistical perspective. These learning experiences are immersion into data-rich environments, probability modeling, an emphasis on using visualizations for conceptual development, a focus on evaluating data-based arguments, and fostering statistical reasoning. We also argue that these curriculum approaches cannot be embedded and implemented without attention to professional development of teachers and assessment practices. New research orientations emanating from these possible changes are identified.

Keywords Essential statistical learning experiences • Data-rich environments • Probability modeling • Visualizations • Designing investigations • Evaluating statistical arguments • Fostering statistical argumentation

12.1 Imagining Statistics Curricula in the Twenty-First Century

Continuing rapid transformations and innovations in our society lead us to constantly evaluate and reimagine curricular approaches in statistics. Big data are on the horizon, interactive nontraditional statistical graphs are proliferating, and dealing with risk information in everyday life is becoming more prevalent, yet none of these statistical activities are common in current curricula. Nor do we know how curricula might incorporate them into teaching and learning programs. What we do know is that statistics curricula constantly change. Because statistics is a living,

M. Pfannkuch (✉)
Department of Statistics, The University of Auckland, Auckland, New Zealand
e-mail: m.pfannkuch@auckland.ac.nz

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_12

evolving field, change in curricula may reflect an update to contemporary statistical methods, or change can be in response to research findings that change our perspective on how students learn statistics or to lobbying by particular groups. Technological changes allow different learning goals to be sought, different statistical methods to be taught, and different pedagogies to be used. The drivers of innovation and change can be at a global level where technology is revolutionizing the way people interact with the world or can be one person whose message resonates with a wider group who enact and implement the vision enunciated. Other drivers of change are new societal perspectives such as requiring evidence-based argumentation. All of these changes challenge us to rethink essential statistics curriculum learning experiences. Whether the changes are viewed as progressive or retrogressive, curricula are shaped, challenged, and buffeted by people and the learning tools available to them.

Prior to the 1990s, statistics was commonly a small part of school mathematics curricula and was mainly limited to descriptive statistics such as computing the mean and constructing graphs (see more in Chap. 2, this Volume). In the 1990s statistics education research and curricula started to flourish partly in response to publications from august bodies (e.g., National Council of Teachers of Mathematics, American Statistical Association, International Association for Statistical Education), governmental awareness that their citizens needed to be data literate (e.g., Ministério da Educação, 2006), and exhortations from renowned statisticians to reform statistics curricula (e.g., Cobb, 2007; Moore, 1990). In the 2000s statistics became an important part of curricula in many countries including Brazil (Ministério da Educação, 2006), the USA (Advanced Placement Statistics (College Board, 2010); Common Core State Standards Initiative, 2010; Guidelines for Assessment and Instruction in Statistics Education (GAISE), 2007), Germany (Kultusministerkonferenz, 2004a, 2004b, 2012), and New Zealand (Ministry of Education, 2007). These initiatives have started to reform statistics curricula from descriptive statistics to the active exploration of data and probability modeling and to invoke a new paradigm of using a statistical approach rather than a mathematical approach to statistics and probability.

These reforms along with research, which was uncovering how students were reasoning from data, opened the doors to consider that statistics had its own unique ways of thinking (Wild & Pfannkuch, 1999) and consequently required a different approach to teaching than mathematics. Hence, characterizing statistics as an intellectual discipline in its own right with its own ways of thinking and arguing (Moore, 1990) led to imagining new paradigms and the challenging of curricular developers to travel down untrodden paths (Cobb, 2007).

This chapter is focused on reimagining curriculum approaches in statistics and other disciplines including the highlighting of gaps in the research knowledge base. In Sect. 12.2 we discuss how we define the term curriculum, together with the major premises underpinning our thoughts when reimagining curriculum approaches. In Sect. 12.3 we discuss the types of learning experiences that may be operationalized in future curricula, while in Sect. 12.4 attention is given to fostering statistical reasoning. Section 12.5 highlights the cascade of subsequent changes that may be activated when curricula change. The chapter ends with a reflection on the possible impacts on statistics curricula and statistics education research in the future.

12.2 Defining Our Approach to Curriculum

Multiple meanings for the term curriculum have developed as research on teaching and the curriculum have proliferated. Curriculum can refer to the national curriculum document, curriculum materials designed for use by teachers or the curriculum enacted in the classroom. The teacher is pivotal in transforming curriculum tasks or what needs to be learned into how it is learned. Teachers' beliefs about the nature of statistics, their goals for student learning, the learning environment they establish, and their own statistical knowledge influence what and how students learn (Stein, Remillard, & Smith, 2007). In this chapter we focus on what and how students may learn in imagined future curricula, while Chaps. 10 and 16 discuss factors influencing the learning of statistics.

When reimagining the statistics curriculum, we considered many questions to determine what might be important to prepare students for their most likely futures. For example: What learning experiences will prepare students to deal with complexity and ambiguity, to be statistically literate citizens, and to challenge statistically based arguments? What thinking, concepts, and patterns of reasoning are essential to provide cognitive infrastructure that will endure despite rapidly changing technological tools?

Our vision for future curricula is one that aspires to be broad in concept and constantly changing in acknowledgment that statistics education must reflect the constant evolution of statistical knowledge and practice and societal goals. Our chapter is based on the following three premises for future curricula:

12.2.1 Promote Essential Statistical Experiences

In an active learning and experiential learning environment, our premise is that future statistics education curricula may attend to the following three features: the whole statistical inquiry cycle from problem to conclusion (Wild & Pfannkuch, 1999), building and exploring probability models (Konold & Kazak, 2008), and critically evaluating data-based arguments from diverse media (Gal, 2002; Gigerenzer, 2014). Within these three learning experiences, we suggest that curricular developers will be identifying basic knowledge and conceptual building blocks of statistics that they perceive will endure as technology changes (e.g., variation, distribution, randomness, number sense, and graph comprehension) including the development of the language of statistics. Moreover, we envisage curricular developers will pay attention to curriculum coherence in terms of concept development and sequencing of topics "which aim for coherence from a student perspective"

(Bakker & Derry, 2011, p. 6). The emphasis on coherent researched curricula pathways to scaffold students' learning from novice to expert is an area that may take primacy in future curricular development.

12.2.2 Foster Statistical Reasoning

Fostering statistical reasoning includes understanding and researching the context of data (Cobb & Moore, 1997), interrogating the data (Wild & Pfannkuch, 1999), and using the data as evidence for making and supporting claims for both inquiry and advocacy (Gal, 2002). Thinking, reasoning, and arguing from and with data, however, have implications for paying attention to general literacy, for verbalizing and writing cogent arguments, and on learning how to argue in the statistics arena (e.g., Fielding-Wells & Makar, 2015). Furthermore, the proliferation of data-based evidence throughout every discipline suggests that teachers in all disciplines might need to be fully accomplished in teaching students how to argue with data in their discipline (Usiskin, 2014).

12.2.3 Assume Technology is an Integral Part of Statistics Curricula (Cf. Bates & Usiskin, 2016)

Technology has the power to give students access to previously inaccessible concepts and ideas, to explore statistics and probability ideas, and to promote a deeper level of understanding of statistics (e.g., Shaughnessy, 2007). Through using technology students begin to think in new ways (e.g., visually), restructure their thought processes, and cognitively stimulate new conceptual infrastructure (e.g., Garfield, delMas, & Zieffler, 2012; Konold & Higgins, 2003). Because technology is constantly evolving, statistical ideas cannot be dependent on specific technology. The enduring ideas of statistical knowledge, thinking, reasoning, and arguing need to be transferred as technology changes as well as conceptualized, identified, and developed in students.

Our premises with regard to reimagining curriculum approaches we believe are essential, if statistics learning is to progress and to be in concert with current statistical practice and thinking and with the needs of a society that is becoming more dependent on data-based evidence and learning from data.

12.3 Essential Learning Experiences

Statistics curricula expectations have gradually changed over the last 30 years to emphasize not only the *content* knowledge to be learned but also the involvement of students in the experience of *doing* statistics through investigations and the importance of developing their statistical *thinking* (e.g., GAISE, 2007). We envisage all three aspects will continue to be the foundations on which curricula will be built. What may change is our current conception of what statistics is (see Chaps. 1 and 4). In this section, however, we discuss how essential learning experiences can be promoted through the provision of data-rich environments for inquiry, probability modeling, emphasis on using visualizations for enduring conceptual development, and attention to designing investigations. For statistical literacy we discuss the need to promote students' abilities in evaluating data-based arguments.

12.3.1 Data-Rich Environments

With the advent of statistics software designed specifically for learning (e.g., *Fathom* (Finzer, 2005), *TinkerPlots* (Konold & Miller, 2011), and *iNZight* (Wild, 2012)) and the American Statistical Association promoting "more data and concepts, less theory and fewer recipes" (Franklin & Garfield, 2006), data-rich learning environments are flourishing at all curricular levels. These technologies have stimulated a rich repository of research (e.g., CATALST, 2012; Connections Project, 2007) on students' statistical thinking and reasoning processes when engaging in investigations (see Chaps. 5 and 8). Data-rich environments involve students' engagement with understanding the context of situations and the questions that need to be and can be answered with data, determining and debating the measures and design used or to be used, analyzing and interrogating multivariate data sets, unlocking stories in the data, and communicating and evaluating findings. Data is authentic, allowing students to experience the uncertainties associated with finding patterns and relationships. Powerful visualizations are also beginning to play an important role in revealing stories in the data and supporting data-based arguments (see Rosling, 2010). Furthermore, statistical investigations and exploratory data analysis are at the heart of some school and tertiary curricula. Therefore, in reimagining curriculum approaches, we envisage that engaging with authentic data will become a cornerstone of curricula with possible steps into big data at school level but definitely at the tertiary level (see Chaps. 13 and 15).

12.3.2 Probability Modeling

Teaching probability lags behind teaching statistics, and worldwide there is general consternation among educators that school curricula are de-emphasizing probability and the development of probabilistic thinking (e.g., Mooney, Langrall, & Hertel, 2014). Although probability is difficult to teach and many misconceptions among students and adults have been identified (e.g., Saldanha & Liu, 2014; Shaughnessy, 2003), we live in a world permeated by chance phenomena and risk that increasingly demands that students understand chance arguments. As Pratt (2011, p. 891–892) stated:

> The statistics curriculum is responding to approaches offered by EDA [exploratory data analysis], new technology, and an understanding of IIR [informal inferential reasoning]. Yet the probability curriculum is not changing. As a result, while the teaching and learning of statistics takes on an enquiry-based problem-solving stance, where students act as data detectives, the pedagogy of probability is ever more isolated in its strange world of coins, spinners and dice as tools for demonstrating in a rough and ready way the existence of theoretical probability.

We contend that the teaching of probability in many countries reflects this sentiment. Many education researchers (e.g., Fielding-Wells & Makar, 2015; Prodromou, 2014) agree with Pratt (2011, p. 897) that probability conceptions should "develop around the notion of probability as a modeling tool that could be used to build models in computer-based simulations akin to video-games that engage children and adolescents of today." This does not mean that coin, dice, and spinner scenarios, which underpin probability, will be lost, rather students can build models of these scenarios using technology such as the innovative software *TinkerPlots* (English & Watson, 2016). Through this technology students can also experience and model real-world scenarios (Konold & Kazak, 2008).

To improve and to study students' probabilistic conceptions, researchers have developed a number of software tools (e.g., *Probability Explorer* (Lee & Lee, 2009), *Basketball Simulation* (Prodromou, 2014)) that allow students to explore the behavior of pre-built probability models. Students typically explore the consequences of actions and conditions such as varying input parameters and observing the resultant outputs. They give students opportunities to test hypotheses, pose "what if" questions, and reason about relationships between the changes and the outcomes. These types of tools appear to be successful in enhancing students' probabilistic notions and conceptual development. More development and research would be invaluable at all curricular levels in this area including how successful small-scale research projects can be replicated on a much larger scale (Biehler, Ben-Zvi, Bakker, & Makar, 2013). It is urgent that better ways are found to develop students' probabilistic thinking and to develop a probability curriculum that is more inquiry-based.

We believe, however, that *Tinkerplots*, which allows students to *construct* models and then explore the behavior of the models they have built, provides a good example for future software development. A body of research on students' ability to construct models using *TinkerPlots* and the consequent development of their probabilistic reasoning is being built up (see Chap. 7). Unlike simulations where variables are prescribed, constructing models involves students in the development of measures and attributes and sequencing events to describe and answer questions about a system. Capturing relevant elements in a model that mimics the random behavior of a system is an essential probability modeling experience. Hence in our reimagined curriculum approaches, the affordances offered by *Tinkerplots* type software will be an integral part of the probability curriculum, a curriculum that no longer separates probability and statistics but one that strongly connects the two together. The future approach to probability is predicted to be through modeling (Chaput, Girard, & Henry, 2011; Eichler & Vogel, 2014; Jones, 2005), an approach to probability that is more aligned with the practice of today's applied probabilists (Pfannkuch & Ziedins, 2014). Thus, we recommend that probability modeling be considered an important component of new curricula and an area to be addressed in research.

12.3.3 Visualizations for Conceptual Development

With the proliferation of research into students' reasoning, it is now possible to design curricula that are based on how reasoning develops in students rather than being based on beliefs about how novices develop understanding about expert practice (Garfield, Le, Zieffler, & Ben-Zvi, 2015; Konold, 2002). New developments in technology have resulted in the ability to use simulations and dynamic visualizations to target and reveal concepts that were previously inaccessible to students. In our reimagined curriculum approaches, the focus will be on development of key concepts including a mapping of the progression of conceptual development across year levels. Some work has started on conceptual development for some aspects of statistics and probability, but much more research is needed on identifying and constructing conceptual pathways (see Chap. 16). We now describe some recent developments in order to illustrate how research can inform curriculum design and the potential power of visualizations, hands-on or computer-based, in enabling better conceptual development.

Bakker, Biehler, and Konold (2005) questioned curricula that introduced middle school students to box plots. Their analysis of the requirements for interpreting box plots highlighted that students' thinking needed to be transitioned from their prior experience of individual cases to the aggregate (Konold, Higgins, Russell, & Khalil, 2015) and from frequency displays to density displays. Furthermore, students viewed the box plot median as a cut point rather than a distribution property, and their sense of the quartile divisions did not lead to notions of measures of spread. When comparing box plots, the students tended to compare the five summary numbers, and when all were higher, they "would conclude that one group had 'larger values' than the other [but] when these differences were not all in the same direction, they did not know what to conclude" (p. 170), a finding that contributed to the idea of informal inference (see Chap. 8). Bakker et al. (2005) concluded that box plots should not be in the curriculum until at least secondary school and when introduced sufficient instructional time should be allocated. They also believed the procedures of how to find the five-number summary and constructing the box plot were counterproductive to conceptual understanding.

The visual tools in *TinkerPlots* seem to provide a possible conceptual pathway from dot plots to box plots. Allmond and Makar (2014) describe how 12-year-olds in a problem context started with the intuitive visual tools of *TinkerPlots* to divide the dot plots of the data into three parts, moved to visual hat plots (a precursor to a box plot), and then standardized hat plots with their measures of center and spread. The final challenge for students was to construct a visual tool to combine the information, which Allmond and Makar claim "resulted in a fairly seamless transition to the box plot" (p. 6). In this non-computational approach, they believed the students could visualize the distributions, the aggregate, and understood the purpose of the median and IQR, but they did not mention whether the students had grasped the underlying concept of density. Hence through an emphasis on visualizations, it would seem that students could access some of the underpinning concepts about comparing groups through developing intuitions from the concrete to the abstract.

Martignon and Krauss (2009) were inspired by research on the decision-making of physicians and investment managers (e.g., Gigerenzer, 2002) and the work of Fischbein (1975) on students' probabilistic intuitions to challenge the notion that conditional probability logic should be taught at the secondary level. Based on the conviction that "stochastic literacy is a necessary condition for enlightened decision-making" (p. 117), they noted that current curricula were not preparing young students to understand how probabilities shape judgments and decision-making in an information-based society. Therefore, they developed a sequence of hands-on activities with concrete materials that enabled 11-year-old students to visualize conditional probability situations. They described how they scaffolded students from deterministic decision-making to probabilistic decision-making practices. Similar to the notion that students are initially better able to think with individual cases, they worked with representations such as tinker cubes and tinker towers "where tangible units encode not just individuals but their features" (p. 138). Their research indicated that enhancing students' perceptual capacity in the areas of proportional and probabilistic reasoning was a necessary step "towards probabilistic comparisons for decision-making and reckoning with risk" (p. 144). They stated unequivocally that if students are to reach competency within the probability arena, then it is essential that probabilistic reasoning and ideas about risk should be stimulated and be in the curriculum before the age of 10.

In a similar more informal vein and using *TinkerPlot*s technology, Konold, Harradine, and Kazak (2007) described how students engaged with conditional probability ideas as they learned to create data factories that could produce imagined objects such as cats, candies, and skateboards. Pratt and Noss (2010), on the other hand, described how *ChanceMaker* technology was designed deliberately to move students toward abstraction of concepts. Their tasks involved students testing conjectures and mending gadgets to facilitate conceptualization of ideas about fairness and randomness. All these studies that we have described involved play-based activities where young students could be introduced to key concepts in probability through visualizations. The curricular approach is to plant the seeds of concepts early in order for the concepts to be developed progressively toward more abstraction and more connectivity with other concepts.

Another area where visualizations for conceptual development have flourished is in the arena of statistical inference. With the realization that introductory students were finding statistical inference difficult and that the myriad of underpinning concepts behind inference were not being established informally at earlier levels, the idea of informal inference gained traction in research (see Chap. 8) and consequently in some curricula designs (e.g., Garfield et al., 2012; Wild, Pfannkuch, Regan, & Horton, 2011). Focusing on the foundations of inferential reasoning not only led to curricula that attended to conceptual development using visualizations across the year levels but also to a reassessment of current inference practices resulting in these curricula adopting randomization and bootstrapping approaches (e.g., Garfield et al., 2012; Madden, 2008; Parsonage, Pfannkuch, Wild, & Aloisio, 2016; Pfannkuch, Budgett, & Arnold, 2015; Tintle, Topliff, Vanderstoep, Holmes, & Swanson, 2012). It is noteworthy that the adoption of the randomization and bootstrapping approaches in these curricula involves using technology that allows conceptual development through enabling students to visualize a phenomenon, to analyze directly the behavior of the phenomenon, and to visualize statistical processes in ways that were not previously possible, such as viewing a process as it develops over time rather than analyzing it from the end result.

According to Garfield et al. (2015, p. 339), the curriculum should be organized "to recognize meaningful patterns of knowledge (e.g., web of concepts, visualization of key concepts, and the relations among them)," a premise that we concur with in our reimagined curriculum approaches. More research on how and what type of visualizations can improve statistical conceptual understanding is now essential as well as identifying key concepts and mapping their development across the curriculum levels.

12.3.4 Designing Investigations

Some school (e.g., Ministry of Education, 2007) and tertiary (e.g., MacGillivray & Pereira-Mendoza, 2011) curricula emphasize the importance of students experiencing the whole statistical inquiry cycle from problem to conclusion. Students are encouraged to do projects where they have the opportunity to pose a question, design an experiment or survey, and collect and analyze their data. In practice much of the learning is focused on students using pre-existing data with the emphasis on the analysis and interpretation of the data. Within the constraints of a classroom or course, projects consume time and resources and do not allow for student-interest problems to be engaged in, such as allocating drugs to participants or observing the effects of sleep deprivation (Bulmer & Haladyn, 2011). Furthermore, one project, while an invaluable experience, does not give students the multiple experiences that would be beneficial in providing them with an array of problem scenarios to draw upon when confronted with a new situation. Nor are they given opportunities to develop the many statistical concepts associated with linking the study design and the analysis and realizing how the design can affect outcomes.

Among researchers and teachers, therefore, is a belief that students' lack of understanding about study design is affecting their ability to acquire the thinking and practice of statisticians. Some researchers (e.g., Konold & Higgins, 2002; Watson & English, 2015) have sought to address this problem through involving young students in thinking about posing investigative and survey questions and collecting data from students in their class. However, there is limited research on students' conducting entire investigations or attending to study design (see Chap. 4).

At the tertiary level, some educators (e.g., Baglin, Bedford, & Bulmer, 2013; Bulmer & Haladyn, 2011; Darius, Portier, & Schrevens, 2007; Steiner & MacKay, 2009) recognized that engaging students with the up-front phases of statistical investigation was necessary. All had experienced problems with projects, particularly with large classes, with toy experiments that did not result in transfer of learning to real-world experiments, and with students' inabilities to connect design and data production to analysis. To improve students' design skills, these educators provided students with a design problem and context through the creation of virtual environments where the purpose was for the students to experience study design and data collection, that is, the obtaining of data. As Steiner and MacKay (2009, p. 364) stated, their virtual environment forced students to "repeatedly answer important questions … What is the goal of the investigation? Should we use an observational or experimental plan? What sampling protocol should we use? What inputs/outputs should we measure or set?" To these questions Darius et al. (2007) add that fundamental to design is context, which includes background knowledge about the environment. These educators reimagined curriculum approaches in response to "the holes in our educational fabric … where methodology meets context" (Wild, 2007, p. 225) through attending to development of essential statistical concepts for statistical reasoning in designing investigations through a virtual environment learning strategy.

An example of a virtual environment is that of Bulmer and Haladyn (2011) who designed an *Island* with virtual inhabitants on whom the students can conduct a wide variety of experiments, observational studies, and surveys. The data are collected in collapsed real time, and since the *Island* has been inhabited since 1779, students have access to ancestral health and demographic records. Based on actual research data such as effects of alcohol on blood pressure, body temperature, and general health, simulations are used to generate the data that students observe. Students need to pose a question; design their study; decide what measurements to take; contact a sample of inhabitants, who can refuse to participate; apply the treatments and tasks; and collect the data. Once the data are collected, they can transfer the data to a statistical program to analyze the data. Another example of a virtual environment is that of Darius et al. (2007) who designed a greenhouse applet where the goal is to find the optimal dose of nitrogen fertilizer that produces a maximum average biomass for 144 tomato plants. The greenhouse has heating elements and lights, both of which affect plant growth. Students need to place each tomato plant on the greenhouse floor, define and assign treatments, define and assign blocking factors and levels, and decide how long plants should grow. Once the design is in place, the students press the grow button and the growth of each plant is then simulated. The students then analyze the data produced.

According to these educators, the virtual environments enhance student engagement (Baglin et al., 2013), allow the flexibility for mistakes and ability to redo experiments (Steiner & MacKay, 2009), and give opportunities to enhance student communication and argumentation. Students also perceive that their learning is improved (Baglin et al., 2013). Darius et al. (2007) note that because students had different experimental designs for the greenhouse, they could compare, contrast, and discuss principles of design more effectively during teaching. Furthermore, the greenhouse acted as a touchstone example for student understanding when they illustrated more complex experimental designs. Although there is no research evidence, these educators believe that the use of virtual environments enables students to grasp more fully the ideas behind study design and the linkages between design and analysis.

Wild (2007) described virtual environments as having unparalleled potential to augment statistical reasoning and thinking and believed they would develop further if they were modular and open source. Virtual environments lack research on student interaction with them and how the environments might be enhancing students' statistical thinking. It is a rich area for curriculum consideration and research. The virtual environments that these educators created have the following hallmarks, which future research could seek to emulate: access to previously inaccessible concepts and ideas, identification of a gap in student understanding about design issues that is at the level of an enduring idea, promotion of essential learning experiences, fostering of statistical argumentation, and closing of the gap between education and statistical practice. Therefore, in our reimagined curriculum approaches, students will be able to enter virtual environments that "will be surrogates for the 'real world'; virtual worlds in which students can design and conduct investigations" (Wild, 2007, p. 323).

12.3.5 Evaluating Arguments and Statistical Literacy

Why can students "compute a standard deviation [yet cannot spot an] example of poor statistical reasoning" (Utts, 2010, p. 1)? Why do students not distinguish between "absolute and relative risks" in media stories (Kurz-Milcke, Gigerenzer, & Martignon, 2008, p. 18)? Educators and researchers are becoming increasingly worried that students are not learning statistical ideas that are needed to make informed decisions in daily life and about societal issues. Current curricula are inadequate for educating people to evaluate data-based arguments critically (Gal, 2002), to understand risk (Gal, 2005; Gigerenzer, 2014), and to recognize the pitfalls of using heuristics when making judgments (Kahneman, 2011). Gigerenzer (2014) even argues for a revolution in the school curriculum with respect to risk literacy.

Garfield and Ben-Zvi (2008) frame statistical literacy in terms of both data producer and data consumer stating it involves understanding and using basic statistical skills, such as being able to construct, recognize, and interpret displays. From a data consumer perspective, Gal (2002) defines statistical literacy as people's ability to interpret and evaluate critically statistically based information from a wide range of sources and to formulate and communicate a reasoned opinion on such information. Watson (2013) concurs with this definition but argues that it applies to adults, and for students to become statistically literate, they must experience the processes of conducting statistical investigations to be able to judge the claims of others. We claim that all students will remain statistically illiterate if critically evaluating other people's statistically based reports is not explicitly taught (Gal, 2002; Gigerenzer, 2014; Schield, 2010; Utts, 2010).

To become statistically literate, Gal (2002, p. 3) argues that cognitively there needs to be a joint activation of "a knowledge component (comprised of five cognitive elements: literacy skills, statistical knowledge, mathematical knowledge, context knowledge, and critical questions) and a dispositional component (comprised of two elements: critical stance and beliefs and attitudes)." In Gigerenzer's (2014, p. 247) view, risk literacy requires statistical thinking, rules of thumb for "making good decisions in an uncertain world," and learning about the psychology of risk concerning "the emotional and social forces that guide our behavior." To envision including statistical literacy in the curriculum seems daunting considering the complexity of integrating a web of ideas and concepts needed for understanding and critically evaluating statistical evidence in reports. Nevertheless there is research available that looks at the type and levels of students' statistical literacy and ways statistical literacy might be conceived and practiced in future curricula.

Watson (1997) researched school students' interpretation of media reports and from the student data developed a hierarchy of three levels necessary for statistical literacy: basic understanding of probabilistic and statistical terminology, understanding of statistical language and concepts embedded in wider social discussion, and challenging claims in the media. Watson and Callingham (2003, p. 20) built on this hierarchy and found that "statistical literacy is a complex construct that may be thought of as a thick thread or rope comprising two interwoven essential strands: mathematical/statistical understanding of the content and engagement with the context in exploiting this understanding." Furthermore, Watson (2013, p. 60) suggested that as students move through the school curriculum experiencing the issues and uncertainties associated with statistical investigations, they should "be exposed to media claims to test their critical thinking skills." She demonstrated how middle school students could learn how to develop a questioning attitude to claims (Watson, 2008). To test a media claim that brown-eyed people had faster reaction times than people with other eye colors, students could collect data using their class and then augment the data collected from other students. Consequently students could start to appreciate the role of sample size and to develop the propensity to think critically when faced with media claims. Constant exposure to evaluating data-based media arguments that are prevalent in students' everyday lives could be embedded into the school curriculum and attended to in teaching.

Drawing on Gal's (2002) ideas, the following researchers deliberately attended to activating cognitively his prescribed knowledge and dispositional components, directly teaching statistical literacy involving media reports, and exploring students' resultant reasoning. Merriman (2006) found that her 14-year-old students became more aware and skeptical of statistics found in media reports and had to use a much higher level of thinking. Sharma, Doyle, Shandil, and Talakia'atu (2011) found that 13-year-old students could develop critical thinking skills including questioning and challenging data and that context and literacy placed huge demands on students, which teachers were able to ameliorate. Rose (2012) developed a learning trajectory for 17-year-old students in preparation for the introduction of a curriculum standard entitled *Evaluate Statistically Based Reports* and concluded the trajectory needed 12 key components, some of which were literacy support, extending students' current statistical and contextual knowledge base, the use of "worry" questions, and development of students' ability to communicate in writing a critical evaluation of a media report. Budgett and Pfannkuch (2010a, 2010b) interviewed students who had completed an undergraduate course involving the evaluation of data-based media arguments. The students reported and demonstrated in tasks given to them that they had increased their awareness of issues underpinning statistically based information in the media and in everyday life.

For learning about risk argumentation, the most extensive research has been conducted by Gerd Gigerenzer and his associated researchers. For example, Kurz-Milcke et al. (2008), Gigerenzer (2014), and Martignon and Krauss (2009) report that risk literacy can be taught to all ages provided new teaching approaches are employed, such as using natural frequencies rather than probabilities, population diagrams for displaying false positives, and using icons to encode information. Pratt, Levinson, Kent, Yogui, and Kapadia (2012) further developed the idea of risk literacy when they delved into how teachers weighed up evidence when confronted with many sources of evidence for determining whether to go ahead with an operation that could cure a painful spinal condition. Their findings demonstrated the richness and complexity of decision-making in the presence of uncertainty and how personal factors were involved. Hence we believe the probability curriculum must include developing students' understandings about risk, as risk ideas permeate our society and are fundamental in decision-making.

Developing students' propensity to identify statistical situations embedded in everyday contexts and to evaluate critically and challenge data-based arguments should be given high priority in research. "Critical thinking using data is an increasingly important core life skill" (Nicholson, Ridgway & McCusker, 2010, p. 5) and is essential for a vision of society as a "participatory democracy" (Gigerenzer, 2014, p. 261). We believe that the evaluation of data-based arguments must be in the spotlight when reimagining curriculum approaches.

12.4 Fostering Statistical Reasoning

With technology facilitating a refocusing of learning on reasoning from plots rather than constructing plots, consideration needs to be given to fostering the ability of students to reason and argue in the statistics arena. In this section we discuss some current and future issues with regard to argumentation in statistics and other disciplines.

12.4.1 Argumentation in Statistics

Increased emphasis in statistics on interpretation, argumentation, and communication means reasoning from data is paramount, yet it has been identified as problematic. Biehler (1997, p. 176) noticed the problem of verbalization when he stated:

> An adequate verbalization is difficult to achieve and the precise wording is often critical. There are profound problems to overcome in interpreting and verbally describing statistical graphs and tables that are related to the limited expressability of complex quantitative relations by means of a common language.

Ridgway, Nicholson, and McCusker (2007) identified the same problem in their research when they reported that teachers found interpretation the most difficult part to model for their students. They referred to a "scarcity of resources which offer advice on strategies for interpreting data, or on a suitable language to use in describing patterns in data" (p. 1). They believed students needed to experience multiple contextual situations not only to unlock the stories in the data but also to communicate and verbalize those stories. Pfannkuch, Regan, Wild, and Horton (2010) also found, when designing exemplars, that it was difficult to verbalize the rich conceptual repertoire underpinning plots and to express clearly and precisely their interpretations and reasoning from data. In New Zealand the introduction of a focus on communicating statistical reasoning and evidence for claims in the curriculum and assessment resulted in mathematics teachers realizing that they needed to improve students' verbalizations and their general literacy. Hence research is needed on what constitutes good argumentation practices and how to grow and scaffold students' argumentation. Within this argumentation contextual knowledge is essential as data and context are inextricably linked (Cobb & Moore, 1997).

Fostering statistical argumentation should be an integral part of the classroom culture across all levels. Interrogation, argumentation, and reasoning from data need to start as soon as students are introduced to data. Fielding-Wells and Makar (2015) illustrate vividly how inquiry-based statistical classrooms can engage young students in using evidence to back or challenge a claim, while Makar, Bakker, and Ben-Zvi (2016) describe how argumentation-based inquiry norms can be established in a classroom (see Chap. 16 for discussion on creating learning environments where argumentation-based inquiry is central). In our reimagined curriculum approaches, fostering statistical argumentation and reasoning will be paramount including developing ways of transforming the argumentation and language of novice-invented descriptions of evidence to the precise wording and language of experts.

12.4.2 Statistics Argumentation in Other Disciplines

Data science is prevalent in an ever-increasing number of disciplines, which Finzer (2013) says involves mathematics and statistics, discipline context knowledge, and computing and data skills. The rapid and exponential rise of data science that transcends disciplines and subject matter content has resulted in a worrying gap between the “need for a data savvy citizenry” and education practice (Finzer, 2013, p. 1). He argues that all learners need to acquire data habits of mind, such as graph the data and look for and tell the story behind the data, across all the subjects they study. Finzer (The Concord Consortium, 2015) is now leading the development of open source software to serve curriculum development projects with the goal of growing data-literate citizens through facilitating students’ engagement in data exploration and argumentation in many different subject areas. We predict that in the far future, teachers in all disciplines will be teaching students how to use data sources to argue in their subject area because when “learning these other subjects statistics is necessary” (Usiskin, 2014, p. 11).

In our reimagined curriculum approaches, we conjecture that statistics educators and statisticians will need to attend to data science across many fields (see Chaps. 1, 13 and 15). Already in tertiary curricula courses, such as psychology, instructors teach the statistics pertaining to their field within their context (e.g., Rowe, McKinney, & Wood, 2010). Primary and middle school teachers tend to teach across subject areas allowing scope for using data as evidence across many contexts (e.g., Konold & Higgins, 2003). At secondary school, where subjects tend to exist in silos, the use of data in other fields becomes problematic. For example, Jowsey (2007) found that biology teachers tended to use a black box approach to statistics leading to misuse of regression and correlation ideas. Collaboration with statistics teachers could have led to students using the statistical skills they already had (e.g., box plots) rather than regression with which they were not familiar.

In the near future, however, we envisage more attempts at learning how to argue with data in different disciplines. To illustrate how future approaches might begin, we describe how researchers from three different countries have moved to other discipline areas. Erickson (2012) in the USA describes how he collaborated with a history teacher to facilitate students (17-year-olds) to argue and make claims about a phenomenon that changed over time in US history by using census data from 1900 to 2000. The main focus of the student projects was to use the data to help support the story they were telling as well as to conjecture reasons for the occurrence of the phenomenon. The art of telling stories from data is prominent in the research of Ridgway, Nicholson, and McCusker (2008), who decided that providing resources for other subject areas was the only way in the UK to promote curriculum reform in statistics. In social science classrooms, they trialed resources based on topics such as alcohol use and poverty and used an accessible analytic tool for multivariate data. They showed that 11–14-year-olds "across the attainment range can engage with and understand complex messages in [multivariate] data" (p. 5). Another observation was that the social science teachers, although weak in statistics, were able to facilitate students into gaining substantive insights from the data. Moreover, the students were engaged and able to communicate the stories in the data. In the Netherlands the government response to a lack of connection between statistics and natural sciences was to introduce a new subject Nature, Life and Technology for 16–18-year-olds. Dierdorp, Bakker, Eijkelhof, and van Maanen (2011) described how they designed resources by educationalizing elements of authentic practices. They designed resources for identifying the best training program for athletes to improve their physical condition and for monitoring the height of dykes in order for students to learn more about regression and correlation. They believed their strategy supported students' learning about how to use data in these disciplines as well as improving their knowledge and understanding of statistics (see also Dierdorp, Bakker, Ben-Zvi, & Makar, 2017). Thus the gap between practice and education can start to be actively closed through collaboration between statistics and other discipline educators and promotion of data science within other subject areas.

Statistics education researchers could take the lead and prioritize exploring ways to develop students' statistical argumentation across multiple contextual situations because in reimagining curriculum approaches, we envisage that statistical argumentation will be an integral part of the learning of statistics and other disciplines.

12.5 Developing Curriculum Change Approaches

Reimagining curriculum approaches is an essential precursor for the development of new national curriculum documents. Advocates for reform, however, must not only have a vision for new curricula but also must consider strategies for implementing their vision. Each country will have its own system for developing new national curriculum documents and the consequent implementation of the curriculum. Strategies to instigate change will depend on the education culture in a country, but there will be some commonalities across countries. First is the consideration that changing curricula expectations has enormous flow on effects to many stakeholders in an education enterprise (e.g., assessment developers, parents). Second, to ensure implementation of new curriculum approaches, many collaborations may need to be formed and driven by groups willing to be involved at all stakeholder levels. Policymakers, statisticians, researchers, teachers, students, and software and resource/textbook designers may need to work together on designing, interpreting, and implementing curricula to match the vision for *what* is taught and *how* it is taught. Third, new curriculum approaches and ways of learning and assessment may need to be connected and integrated at the school, state/national standard tests, state/national qualifications, and policy and political levels. Fourth, there may need to be recognition that a new curriculum should be an *educative* curriculum (Stein et al., 2007) whereby teachers may need to learn new content and approaches to learning. Some countries may need to seek ways for facilitating teachers and students to leapfrog from 1960s mathematics curricula into contemporary statistics education practices. Finally, the reimagined curriculum may not match the reality in practice, but without a vision the curriculum cannot evolve.

In this section we briefly discuss some potential issues in developing curriculum change approaches to embed new curricula (see Chap. 16 for a full discussion) and the necessity for assessment to be changed to match new curricula learning goals and intent.

12.5.1 Changing to the Twenty-First Century Statistics Curricula

How can a twenty-first century statistics curriculum be provided to all students? Drawing on the perspectives of researchers from two countries, Kenya and Brazil, we discuss their current approaches to this problem and what lessons can be learned more generally. Both countries have similar problems: they note that their teachers lack content knowledge and confidence, teach prescriptively from textbooks, lack experience with technology, and have a mathematical approach to statistics. Although the problems are similar, the solutions appear to depend on the resources and culture of the community. According to Stern (2013, p. 1), Kenya lacks "an educational culture that embraces and encourages" change, resulting in initiatives making no substantive difference to student learning and outcomes. Also, unlike Brazil, there seems to be no established community of mathematics and statistics education researchers and little professional development for teachers.

In Kenya, Stern (2014) and Manyalla, Mbasu, Stern, and Stern (2014) believe that technology may be the key ingredient in provoking change as long as access is free, it can be in the hands of each student, and teaching resources accompany the technology. Their most successful initiative seems to be the use of Computer-Assisted Statistical Textbooks (CAST) (Stirling, 2005), an idea that seems to be very similar to their current reliance on textbooks and where the teachers may feel more confident and comfortable. Manyalla, et al. (2014) describe how students who used CAST outperformed students who did not use CAST in the national examinations resulting in other teachers being willing to try CAST in their classrooms. Furthermore students using CAST were more engaged in statistics, liked the interactive style, and were able to peer teach themselves such that they leaped beyond the statistical knowledge of their teacher (Zachariah Mbasu, personal communication, July 2012).

In Brazil, an approach to change teachers' attitudes toward statistics and statistics teaching has been through teacher learning communities. The ethos behind such learning communities is that innovation and change for teachers takes long-term commitment and is based on teachers and researchers collaborating to identify problem areas and ways of resolving them. For example, Souza, Lopes, and Pfannkuch (2015) describe a potential model for developing middle school teacher expertise where a community of teachers is exposed to new ideas, collaboratively plans lessons and implements them in their classrooms, reports back to the group, and critically reflects on their practice. Nacarato and Grando (2014) describe a similar model except they used videos of implementations for group reflection.

From the experience of these researchers in these two countries, it seems that a general consideration for all countries is to think about curriculum change as an evolutionary process and that facilitators within the community who work alongside teachers are best placed to implement change. A common factor was facilitators starting with groups of teachers who were willing to trial changes in their classroom that were not too far from current practice. When teachers noticed and experienced their students' engagement and success using new statistics teaching approaches, it seemed that they became more responsive to change. We suggest that reimagining curriculum approaches, as these researchers noted, also requires attention to current resources available, the development of resources, the education culture of the community, the adoption of accessible technology for all, and innovative strategic thinking for enhancing teacher knowledge and practice. With the availability of MOOCs (see Chap. 1) and e-resources such as CAST, groups of students and teachers may even be able to bypass moribund education systems and educate themselves. However, the reigning assessment system may prevent such an action.

12.5.2 Assessment

Assessment drives what is taught and valued. As curriculum goals shift emphasis from computational skills toward deeper conceptual understanding, reasoning from data, and evaluating statistical arguments, different types of assessment methods are essential. As Garfield and Ben-Zvi (2008, p. 66) stated, "assessment should be carefully aligned with the important and valued learning goals." If the learning goals of a curriculum are not aligned with assessment practices, then priority in teaching will be given to the assessment goals. Also assessment tasks convey information about the nature and role of the discipline and what thinking and reasoning are valued. In an analysis of the high-stake statistics assessment in the UK, McCusker, Nicholson, and Ridgway (2010) concluded that the tasks conveyed an impoverished view of statistics. Similarly Callingham (2011) suggested that statistics was portrayed as simple mastery of skills for educational measurement purposes in Australia resulting in a limiting primary curriculum. Ridgway et al. (2008) also noted how assessment limited primary and middle school students to univariate data, whereas in practice the students were quite capable of dealing with multivariate data.

For new types of curricula, important learning goals and guiding principles (Garfield & Franklin, 2011) need to be identified and enacted as the following two examples illustrate. Garfield, delMas, and Zieffler (2010) demonstrate how their valued learning outcomes of literacy, reasoning, and thinking could be assessed in a written examination where computer output is given and responses range from extracting statistical information from plots, reasoning about underlying statistical ideas, to interpreting and critiquing statistical claims in a full written format. They also mention ARTIST, an online database of items that assess these outcomes using three types of format (open-ended, multiple-choice, performance task). These researchers identified the learning outcomes that were important for their introductory course and matched the assessment to those outcomes. Similarly, Budgett and Pfannkuch (2010a) describe their statistical literacy course, where the learning goals were to evaluate statistically based studies; to construct statistically sound statements, graphics, and reports; and to recognize statistical concepts in everyday events. Consequently their assessment is aligned to these goals as students are required, for example, to critically evaluate a study, write a statistically sound newspaper report from a study, and reason statistically about everyday events such as recognizing regression to the mean.

In a technology environment—where computations and representations are automated—interpretation, argumentation, and communication can be given primacy in teaching. This type of reasoning is much more demanding conceptually and requires literacy abilities. With access to more powerful ways of understanding statistics, students also have the opportunity "to appreciate utility: how and why … the statistical idea is useful" (Ainley & Pratt, 2010, p. 2). For example, students can experience graphs as tools for interpretation and analysis rather than tools for displaying the end results. Utility is a facet that Ainley and Pratt (2010, p. 2) believe is missing from many approaches to assessment where both procedural and conceptual knowledge seem to be equated with "how to calculate or represent statistical objects." Ainley, Pratt, and Hansen (2006) think designers of tasks should include the assessment of students' understanding, for example, of the utility of a statistical measure (e.g., mean, spread).

The impact of technology on teaching and learning and on the understanding of students has not generally been matched by developments in assessment. If learning involves using technology for analysis, for conceptual development, and to build probability models, then the same technology should be used in assessment. For example, in New Zealand school-based national assessment standards allow students to conduct investigations using technology (e.g., *iNZight*) including writing a full report, while in the USA the Advanced Placement Statistics examination integrates technology through graphing calculators and including statistical output in some questions (Garfield & Franklin, 2011). Furthermore, Callingham (2011) contends that traditional assessment items could disadvantage students who use technology for learning, as they may have developed a different cognitive infrastructure from those students in a traditional setting. She believes that "the nature of the changes to cognition needs to be identified … [and is an] area that needs further research" (p. 9).

Assessing only the mastery of procedural skills will not provide evidence of the new curricular goals of literacy, reasoning, and thinking (Garfield & Ben-Zvi, 2008; Garfield et al., 2010). As Callingham (2011, p. 9) states, "it is time to reconsider the assessment challenge and develop new approaches that take account of both technology use by students and the power of technology to deliver assessment." In our reimagined curriculum approaches, the curriculum should drive assessment, whereby student assessment is aligned with the goals of learning and the technology used by the students. Furthermore, guiding principles and goals (see Garfield & Franklin, 2011) should be developed, implemented, and used by assessment developers whether high stakes or teacher designed.

12.6 Conclusion

It is predicted that change in the first quarter of the twenty-first century will be equivalent to the change in the entire twentieth century. Curricula usually have a lifetime between 10 and 20 years, and consequently any new curriculum approaches must step into the unknown and be prepared to be constantly innovative and responsive to new unimagined possibilities. As new societal learning goals are formulated and technological tools constantly change the way we interact and think with data, an articulated vision of future curriculum approaches is essential to challenge educators to travel down uncharted pathways.

Our reimagining of curriculum approaches applies across the education spectrum from kindergarten to tertiary levels. We acknowledge the constant change wrought by technology, which is changing our thinking tools. We propose that students be immersed in data-rich environments with more emphasis on probability modeling. Virtual environments need to be considered and developed further in order to assist students to connect study design with analysis and also could be expanded into new territory where virtual worlds could become surrogates for real-world statistics and probability problems. These types of learning environments we envisage can facilitate students' understanding of statistics through enabling them to enter into the playgrounds of statistics and probability. Purpose-built learning technology can also advance students' understanding to unprecedented levels through allowing access to previously inaccessible concepts through the visual senses including dynamic visualizations. With the ability of technology to help build conceptual understanding via visualizations, we envisage curricula will emphasize conceptual development and that coherent conceptual pathway progressions across all levels will be mapped for many areas of statistics and probability and in many diverse ways.

With the rapid changes in technology, it has become important to establish and teach the enduring ideas and concepts that underpin statistical knowledge, thinking, reasoning, and arguing. We believe these enduring ideas and concepts should be part of the essential learning experiences of students, experiences that are active, interactive, and coherent. Furthermore, society is demanding evidence-based arguments, which are proliferating in people's everyday life and at societal levels. To engage in a participatory democracy, it is crucial that statistics curricula respond, and therefore, we believe a learning goal should be evaluating data-based arguments, a skill that needs to be directly taught in order to inculcate the necessary higher levels of thinking needed to be truly statistically literate. Fostering statistical argumentation in statistics and other disciplines is a high priority as data science becomes ubiquitous across many fields of endeavor. The automation of plots and analyses can redirect learning to concentrate on learning how to argue with and from data, with the context, and communicating verbally and in writing well-reasoned claims based on the data for the purpose of advocacy or inquiry. Development of such skills is paramount in today's statistical world.

In reimagining curriculum approaches, we also briefly highlighted the large number of stakeholders influenced by changes in curricula. We believe local and global connectivity can help teachers and education systems in transforming their curricula, where pathways to change may be unique to each community and culture. With regard to assessment, matching curricular goals and the technology used for learning is essential for implementation of future curricula aspirations. Curricular change also cannot occur without the international and local statistics communities forming collaborations themselves among statisticians, educators, and researchers and continuing to be prepared as groups to promote their discipline and to be active in promoting a vision for statistics curricular approaches.

Questioning current practices, inventiveness, pushing the frontiers of possibilities, and cutting-edge research are the hallmarks of much of the research and innovations quoted in this chapter. Our articulated vision of future curricula will only come to fruition if researchers and educators continue to follow these researchers' footsteps and provide further foundations for constructing new learning approaches and new curricula. In reimagining curriculum approaches, we recommend possible avenues for future research:

– *More insight into fostering statistical argumentation* including learning how to make evidence-based claims in data-rich environments and critically evaluating data-based arguments in diverse media from a statistical literacy perspective
– *In-depth studies on probability modeling, risk, and designing investigations* to learn how to scaffold students' reasoning and to identify key issues that need to be addressed before designing curricula in these new areas
– *More insight into how visualizations may enhance student conceptual understanding* in order to understand the advantages, pitfalls, and visual design considerations when students reason from visual representations of stochastic ideas and processes
– *Development of coherent curricula conceptual pathways* and infrastructure that pay attention to and identify enduring notions that will prevail despite changes in technology and will make sense from a student perspective (Bakker & Derry, 2011)

As technology continues to shape the statistics discipline and learning approaches, our wish is that researchers in all countries will challenge current practice and continue to re-envisage curriculum approaches in ways not yet imagined.

Acknowledgments I wish to thank the following people for their very helpful suggestions for improving this chapter: Pip Arnold, James Baglin, Stephanie Budgett, Jill Fielding-Wells, Bill Finzer, Christine Franklin, Jennifer Kaplan, and Leandro de Oliveira Souza.

References

Ainley, J., & Pratt, D. (2010). It's not what you know, it's recognizing the power of what you know: Assessing understanding of utility. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute.

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal, 32*(1), 23–38.

Allmond, S., & Makar, K. (2014). From hat plots to box plots in TinkerPlots: Supporting students to write conclusions which account for variability in data. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Baglin, J., Bedford, A., & Bulmer, M. (2013). Student' experiences and perceptions of using a virtual environment for project-based assessment in an online introductory statistics course. *Technology Innovations in Statistics Education*, *7*(2). Retrieved from http://www.escholarship.org/uc/item/137120mt

Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about boxplots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education (IASE) roundtable, Lund, Sweden 28 June-3 July 2004* (pp. 163–173). Voorburg, The Netherlands: International Statistical Institute.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1–2), 5–26.

Bates, M., & Usiskin, Z. (Eds.). (2016). *Digital curricula in school mathematics*. Charlotte, NC: Information Age Publishing.

Biehler, R. (1997). Students' difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics, International Association for Statistical Education Round Table Conference, Granada 1996* (pp. 169–190). Voorburg, The Netherlands: International Statistical Institute.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–689). New York, NY: Springer.

Budgett, S., & Pfannkuch, M. (2010a). Assessing students' statistical literacy. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 103–122). Chichester, UK: John Wiley and Sons Ltd..

Budgett, S., & Pfannkuch, M. (2010b). Using media reports to promote statistical literacy for non-quantitative majors. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics, Ljubljana, Slovenia, July 2010*. Voorburg, The Netherlands: International Statistical Institute.

Bulmer, M., & Haladyn, J. K. (2011). Life on an island: A simulated population to support student projects in statistics. *Technology Innovations in Statistics Education*, *5*(1). Retrieved from http://www.escholarship.org/uc/item/2q0740hv

Callingham, R. (2011). Assessing statistical understanding in middle schools: Emerging issues in a technology-rich environment. *Technology Innovations in Statistics Education*, *5*(1). Retrieved from http://escholarship.org/uc/item/3qr2p70t

CATALST. (2012). Change agents for teaching and learning statistics. Retrieved from https://github.com/zief0002/Statistical-Thinking

Chaput, B., Girard, J., & Henry, M. (2011). Frequentist approach: Modeling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics–challenges for teaching and teacher education: A joint ICMI/IASE study: The 18th ICMI study* (pp. 85–95). New York, NY: Springer.

Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), 1–15. Retrieved from http://escholarship.org/uc/item/6hb3k0nz

Cobb, G., & Moore, D. (1997). Mathematics, statistics and teaching. *The American Mathematical Monthly, 104*(9), 801–823.

College Board. (2010). *Advanced Placement statistics course description*. Princeton, NJ: College Board. Retrieved from http://media.collegeboard.com/digitalServices/pdf/ap/ap-statistics-course-description.pdf

Common Core State Standards Initiative. (2010). Common Core State Standards for mathematics. In *Common Core State Standards (College- and Career-Readiness Standards and K12 standards in English language arts and math)*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. Retrieved from http://www.corestandards.org

Connections Project. (2007). Retrieved from http://dbz.edtech.haifa.ac.il/developing-statistical-thinking

Ministério da Educação, M. (2006). *Parâmetros curriculares nacionais: Matemática (National curricular parameters: Mathematics)*. Brasilia, Brazil: Author.

Darius, P., Portier, K., & Schrevens, E. (2007). Virtual experiments and their use in teaching experimental design. *International Statistical Review, 75*(3), 281–294.

Dierdorp, A., Bakker, A., Ben-Zvi, D., & Makar, K. (2017). Students' considerations of variability in measurement activities based on authentic practices. *Statistics Education Research Journal*, *16*(2).

Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*(1–2), 132–151.

Eichler, A., & Vogel, M. (2014). Three approaches for modeling situations with randomness. In E. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 75–99). Dordrecht, The Netherlands: Springer.

English, L., & Watson, J. (2016). Development of probabilistic understanding in fourth grade. *Journal for Research in Mathematics Education, 47*(1), 28–62.

Erickson, T. (2012). Signs of change: History revealed in U.S. census data. Retrieved from http://www.amazon.com/Signs-Change-Statistics-Lick-Wilmerding-2011-2012/dp/1479222046

Fielding-Wells, J., & Makar, K. (2015). Inferring to a model: Using inquiry-based argumentation to challenge young children's expectations of equally likely outcomes. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 1–27). Minneapolis, MN: Catalyst Press.

Finzer, W. (2005). Fathom™ dynamic statistics software (Version 2.0). Emeryville, CA: Key Curriculum Press.

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education*, *7*(2). Retrieved from http://escholarship.org/uc/item/7gv0q9dc

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht, The Netherlands: Reidel.

Franklin, C., & Garfield, J. (2006). The GAISE project: Developing statistics education guidelines for grades pre-K-12 and college courses. In G. Burrill & P. Elliott (Eds.), *Thinking and rea-*

soning with data and chance, sixty-eighth yearbook (pp. 335–375). Reston, VA: The National Council of Teachers of Mathematics.

GAISE K-12 Report. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-k-12 curriculum framework*. Alexandria, VA: American Statistical Association.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–51.

Gal, I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 39–63). New York, NY: Kluwer/Springer Academic Publishers.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, The Netherlands: Springer.

Garfield, J., delMas, R., & Zieffler, A. (2010). Assessing important learning outcomes in introductory tertiary courses. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 75–86). Chichester: John Wiley and Sons Ltd..

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM, 44*(7), 883–898.

Garfield, J., & Franklin, C. (2011). Assessment of learning, for learning, and as learning. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics–challenges for teaching and teacher education: A joint ICMI/IASE study: The 18th ICMI study* (pp. 133–145). New York, NY: Springer.

Garfield, J., Le, L., Zieffler, A., & Ben-Zvi, D. (2015). Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking. *Educational Studies in Mathematics, 88*(3), 327–342.

Gigerenzer, G. (2002). *Reckoning with risk: Learning to live with uncertainty*. London: Penguin Books.

Gigerenzer, G. (2014). *Risk savvy: How to make good decisions*. New York, NY: Viking.

Jones, G. (Ed.). (2005). *Exploring probability in school: Challenges for teaching and learning*. New York, NY: Springer.

Jowsey, M. (2007). Statistical thinking across the curriculum: Year 13 biology students. Unpublished Master's Thesis. The University of Auckland, New Zealand.

Kahneman, D. (2011). *Thinking fast and slow*. New York, NY: Farrar, Straus and Giroux.

Konold, C. (2002). Teaching concepts rather than conventions. *New England Journal of Mathematics, 34*(2), 69–81.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematics Learning, 12*(3), 217–230.

Konold, C., & Higgins, T. (2002). Highlights of related research. In S. J. Russell, D. Schifter, & V. Bastable (Eds.), *Developing mathematical ideas: Working with data* (pp. 165–201). Parsippany, NJ: Dale Seymour Publications.

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., Higgins, T., Russell, J., & Khalil, K. (2015). Data seen through different lenses. *Educational Studies in Mathematics, 88*(3), 305–325.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, *2*(1). Retrieved from http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art1/

Konold, C., & Miller, C. (2011). TinkerPlots (Version v2.0) [Computer software]. Emeryville, CA: Key Curriculum Press.

Kultusministerkonferenz. (2004a). *Bildungsstandards im Fach Mathematik für den Primarbereich. Beschluss vom 15.10.2004*. München, Germany: Wolters-Kluwer.

Kultusministerkonferenz. (2004b). *Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss*. München, Germany: Wolters Kluwer.

Kultusministerkonferenz (2012). Bildungsstandards im Fach Mathematik für die allgemeine Hochschulreife. Retrieved from: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Mathe-Abi.pdf

Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication. *Annals of the New York Academy of Sciences, 1128*, 18–28.

Lee, H. S., & Lee, J. T. (2009). Reasoning about probabilistic phenomena: Lessons learned and applied in software design. *Technology Innovations in Statistics Education*, *3*(2). Retrieved from http://www.escholarship.org/uc/item/1b54h9s9

MacGillivray, H., & Pereira-Mendoza, L. (2011). Teaching statistical thinking through investigative projects. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Teaching statistics in school mathematics - challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 109–120). New York, NY: Springer.

Madden, S. (2008). *Dynamic technology scaffolding: A design principle with potential for supporting statistical conceptual understanding*. Monterrey, Mexico: International Congress on Mathematics Education—ICME-11. Retrieved from http://tsg.icme11.org/document/get/479

Makar, K., Bakker, A., & Ben-Zvi, D. (2016). Scaffolding norms of argumentation-based inquiry in a primary mathematics classroom. *ZDM, 47*(7), 1107–1120.

Manyalla, B., Mbasu, Z., Stern, D., & Stern, R. (2014). Measuring the effectiveness of using computer assisted textbooks in Kenya. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Martignon, L., & Krauss, S. (2009). Hands-on activities for fourth graders: A tool box for decision-making and reckoning with risk. *International Electronic Journal of Mathematics Education, 4*(3), 117–148.

McCusker, S., Nicholson, J., & Ridgway, J. (2010). Statistics assessment: The good, the bad, and the ugly. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute.

Merriman, L. (2006). Using media reports to develop statistical literacy in year 10 students. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics (ICOTS7, July, 2006), Salvador, Brazil*. International Statistical Institute: Voorburg, The Netherlands.

Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, New Zealand: Learning Media.

Mooney, E., Langrall, C., & Hertel, J. (2014). A practitional perspective on probabilistic thinking models and frameworks. In E. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 495–508). New York, NY: Springer.

Moore, D. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.

Nacarato, A., & Grando, R. (2014). Teachers' professional development in a stochastics investigation community. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Nicholson, J., Ridgway, J., & McCusker, S. (2010). Luring non-quantitative majors into advanced statistical reasoning (and luring statistics educators into real statistics). In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute.

Parsonage, R., Pfannkuch, M., Wild, C. J., & Aloisio, K. (2016). Bootstrapping confidence intervals. In D. Ben-Zvi & K. Makar (Eds.), *The teaching and learning of statistics* (pp. 181–191). Switzerland: Springer International publishing.

Pfannkuch, M., Budgett, S., & Arnold, P. (2015). Experiment-to-causation inference: Understanding causality in a probabilistic setting. In A. Zieffler & E. Fry (Eds.), *Reasoning about uncertainty: Learning and teaching informal inferential reasoning* (pp. 95–125). Minneapolis, MN: Catalyst Press.

Pfannkuch, M., Regan, M., Wild, C. J., & Horton, N. (2010). Telling data stories: Essential dialogues for comparative reasoning. *Journal of Statistics Education*, *18*(1). Retrieved from http://www.amstat.org/publications/jse/v18n1/pfannkuch.pdf

Pfannkuch, M., & Ziedins, I. (2014). A modeling perspective on probability. In E. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 101–116). New York, NY: Springer.

Pratt, D. (2011). *Re-connecting probability and reasoning from data in secondary school teaching. Proceedings of the 58th International Statistical Institute World Statistical Congress, Dublin* (pp. 890–899). The Hague, The Netherlands: International Statistical Institute.

Pratt, D., Levinson, R., Kent, P., Yogui, C., & Kapadia, R. (2012). A pedagogic appraisal of the priority heuristic. *ZDM, 44*(7), 927–940.

Pratt, D., & Noss, R. (2010). Designing for mathematical abstraction. *International Journal of Computers for Mathematics Learning, 15*, 81–97.

Prodromou, T. (2014). Developing a modeling approach to probability using computer-based simulations. In E. J. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 417–439). Dordrecht, The Netherlands: Springer.

Ridgway, J., Nicholson, J., & McCusker, S. (2007). Embedding statistical assessment within cross-curricular materials. In B. Phillips (Ed.), *Proceedings of the International Association of Statistical Education Satellite Conference on Assessing Student Learning in Statistics, Guimaraes, Portugal, 19–21 August, 2007*. Voorburg, The Netherlands: International Statistical Institute.

Ridgway, J., Nicholson, J., & McCusker, S. (2008). Reconceptualizing 'statistics' and 'education'. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference. Monterrey, Mexico, July 2008*. Retrieved from https://iase-web.org/documents/papers/rt2008/T1P5_Ridgway.pdf

Rose, D. (2012). Developing statistical literacy in year 13 statistics students. Unpublished Master's thesis. The University of Auckland, New Zealand.

Rosling, H. (2010). *What showbiz has to do with it*. Plenary presented at the 8th International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia. Retrieved from http://videolectures.net/icots2010_rosling_wshdw/

Rowe, R., McKinney, P., & Wood, J. (2010). Inquiry-based assessment of statistical methods in psychology. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 189–199). Chichester: Wiley.

Saldanha, L., & Liu, Y. (2014). Challenges of developing coherent probabilistic reasoning: Rethinking randomness and probability from a stochastic perspective. In E. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 367–396). New York, NY: Springer.

Schield, M. (2010). Assessing statistical literacy: Take care. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 133–152). Chichester: Wiley.

Sharma, S., Doyle, P. Shandil, V., & Talakia'atu, S. (2011). Statistics is boring … because it makes you think! Summary research report for the Teaching and Learning Research Initiative. Retrieved from http://www.tlri.org.nz/sites/default/files/projects/9270_sharma-summaryreport.pdf

Shaughnessy, M. (2003). Research on students' understandings of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216–226). Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second handbook of research on the teaching and learning of mathematics* (Vol. 2, pp. 957–1009). Charlotte, NC: Information Age Publishers.

Souza, L., Lopes, C., & Pfannkuch, M. (2015). Collaborative professional development for statistics teaching: A case study of two middle school teachers. *Statistics Education Research Journal, 14*(1), 112–134.

Stein, M., Remillard, J., & Smith, M. (2007). How curriculum influences student learning. In F. Lester (Ed.), *Second handbook of research on the teaching and learning of mathematics* (Vol. 1, pp. 319–369). Charlotte, NC: Information Age Publishers.

Steiner, S., & MacKay, J. (2009). Teaching variation reduction using a virtual manufacturing environment. *The American Statistician, 63*(4), 361–365.

Stern, D. (2013). Developing statistics education in Kenya through technological innovations at all academic levels. *Technology Innovations in Statistics Education*, *7*(2). Retrieved from http://www.escholarship.org/uc/item/5cc0s29n

Stern, D. (2014). Reflections on using technology to teach statistics in Kenya. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Stirling, D. (2005). CAST for Africa. Retrieved from http://cast.massey.ac.nz/collection_african.html

The Concord Consortium. (2015). Common Online Data Analysis Platform. Retrieved 2015, February 22 from http://concord.org/projects/codap

Tintle, N., Topliff, K., Vanderstoep, J., Holmes, V., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal, 11*(1), 21–40.

Usiskin, Z. (2014). On the relationships between statistics and other subjects in the K-12 curriculum. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Utts, J. (2010). Unintentional lies in the media: Don't blame journalists for what we don't teach. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute.

Watson, J. (1997). Assessing statistical thinking using the media. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107–121). Amsterdam, The Netherlands: IOS Press.

Watson, J. (2008). Eye colour and reaction time: An opportunity for critical statistical reasoning. *Australian Mathematics Teacher, 64*(3), 30–40.

Watson, J. (2013). Statistical literacy, a statistics curriculum for school students, the pedagogical content needs of teachers, and the Australian Curriculum. *Curriculum Perspectives, 33*(3), 58–69.

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3–46.

Watson, J., & English, L. (2015). Introducing the practice of statistics: Are we environmentally friendly? *Mathematics Education Research Journal, 27*(4), 585–613.

Wild, C. J. (2007). Virtual environments and the acceleration of experiential learning. *International Statistical Review, 75*(3), 322–335.

Wild, C. J. (2012). iNZight. Retrieved from https://www.stat.auckland.ac.nz/~wild/iNZight/

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review, 67*(3), 223–265.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society, 174*(2), 247–295.

Chapter 13
Challenge to the Established Curriculum: A Collection of Reflections

Robert Gould, Roger D. Peng, Frauke Kreuter, Randall Pruim, Jeff Witmer, and George W. Cobb

Abstract We invited a number of prominent statisticians and statistics educators to glimpse into the future to discuss what they see as the significant challenges to the established statistics curriculum that enculturate students into statistical practices that underpin the activity of statisticians. Peng, Kreuter, and Gould discuss various developments, which are already gaining traction in current society and will support the notion of immersion in a data-rich curriculum. The influence of MOOCs, "big data," and Bayesian approaches is primarily discussed by these writers in relation to an undergraduate curriculum. Pruim raises some key questions about teaching computation in statistics with a particular emphasis on undergraduates and programming. In the final piece of writing, Witmer and Cobb discuss the increasing influence

R. Gould (✉)
Department of Statistics, University of California, Los Angeles, CA, USA
e-mail: rgould@stat.ucla.edu

R.D. Peng
Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA
e-mail: rpeng@jhsph.edu

F. Kreuter
University of Maryland, College Park, MD, USA

University of Mannheim, Mannheim, Germany
e-mail: fkreuter@umd.edu

R. Pruim
Department of Mathematics and Statistics, Calvin College, Grand Rapids, MI, USA
e-mail: rpruim@calvin.edu

J. Witmer
College of Arts and Sciences, Oberlin College, Oberlin, OH, USA
e-mail: jeff.witmer@oberlin.edu

G.W. Cobb
Mt. Holyoke College, Hadley, MA, USA
e-mail: gcobb@mtholyoke.edu

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_13

of Bayesian inference with an emphasis on a curriculum that fosters statistical reasoning and the evaluation of arguments.

Keywords Data analysis • MOOC • Big data • Data science • Technology • Bayesian statistics • Secondary education • Computation • Analytics • Messy data • Participatory sensing • Programming

13.1 Introduction

Chapter 12 (Pfannkuch, this volume) reimagines curriculum approaches that enculturate students into statistical practices that underpin the activity of statisticians but perhaps are not given sufficient emphasis in today's school curriculum. We invited a number of statisticians and statistics educators to glimpse into the future to discuss what they see as the significant challenges to the established curriculum.

In the brief writings presented below, Peng (Sect. 13.2), Kreuter (Sect. 13.3), and Gould (Sect. 13.4) discuss various developments, which are already gaining traction in current society and will support the notion of immersion in a data-rich curriculum as proposed by Pfannkuch in Chap. 12. The influence of MOOCs, "big data," and Bayesian approaches is primarily discussed by these writers in relation to an undergraduate curriculum. We note though that MOOCs offer access to nonspecialists and in high schools most teachers of statistics are in fact nonspecialists in the statistics discipline. Furthermore, the emergence of big data offers opportunities for multidisciplinary work which could be of interest to high schools and demands new ways of thinking about statistical inference as currently taught in high schools.

In her vision, Pfannkuch assumes that technology will be an integral part of statistics curricula. Below, Pruim (Sect. 13.5) raises some key questions about teaching computation in statistics with a particular emphasis on undergraduates and programming. In some countries, for example, the UK, programming has been rediscovered as a key skill for the twenty-first century, and so we believe that the questions raised by Pruim have significance for all levels of schooling.

In the final piece of writing, Witmer (Sect. 13.6) and Cobb (Sect. 13.7) discuss the increasing influence of Bayesian inference, which speaks to Pfannkuch's emphasis on a curriculum that fosters statistical reasoning and the evaluation of arguments.

13.2 The Massive Future of Statistics Education

Roger D. Peng
Department of Biostatistics
Johns Hopkins University
Baltimore, MD, USA
rpeng@jhsph.edu

Everywhere you turn, data are being generated somehow. By the time you read this piece, you'll probably have collected some data. You can't avoid data—it's coming from all directions.

So what do we do with all these data? For the most part, nothing. There's just too much data being spewed about. But for the data that we are interested in, we need to know the appropriate methods for thinking about and analyzing them. And by "we," I mean pretty much everyone.

In the future, everyone will need some data analysis skills. People are constantly confronted with data and the need to make choices and decisions from the raw data they receive. Phones deliver information about traffic, ratings of restaurants or books, and even rankings of hospitals. High school students can obtain complex and rich information about the colleges to which they're applying, while admission committees can get real-time data on applicants' interest in the college.

How will people be trained in statistics and in understanding uncertainty in the future? How can we scale that training to meet that enormous demand that has been generated in such a short period of time? Massive open online courses (MOOCs) offer one possibility to deliver content and training in a high-bandwidth, low-cost format that is accessible to a broad audience.

Our educational system has insufficient capacity. The McKinsey Global Institute, in a highly cited report (Lund, Manyika, Nyquist, Mendonca, & Ramaswamy, 2013), predicted that there would be a shortage of "data geeks" and that by 2018 there would be between 140,000 and 190,000 unfilled positions in data science. In addition, there will be an estimated 1.5 million people in managerial positions who will need to be trained to manage data scientists and to understand the output of data analysis. If history is any guide, it's likely that these positions will get filled by people regardless of whether they are properly trained. The potential consequences are disastrous as untrained analysts interpret complex big data coming from myriad sources of varying quality.

Who will provide the necessary training for these unfilled positions?

The field of statistics' current system of training people and providing them with master's degrees and PhDs is woefully inadequate to the task. In 2013, as reported by the American Statistical Association, the top ten largest statistics master's degree programs in the USA graduated a total of 730 people (Pierson, 2014). At this rate we will never train the people needed. While statisticians have greatly benefited from the sudden and rapid increase in the amount of data flowing around the world, our capacity for scaling up the needed training for analyzing those data is essentially nonexistent.

On top of all this, I believe that the McKinsey report (Lund et al., 2013) is a gross underestimation of how many people will need to be trained in some data analysis skills in the future. Given how many data are being generated every day, and how critical it is for everyone to be able to intelligently interpret these data, I would argue that it's necessary for everyone to have some data analysis skills. Needless to say, it's foolish to suggest that everyone get a master's or even bachelor's degrees in statistics. We need an alternate approach that is both high quality and scalable to a large population over a short period of time.

13.2.1 Massive Open Online Courses (MOOCs)

There has been a major push to create statistical content that can be delivered in the format of a massive open online course (MOOC). A few examples of this include Statistics One by Andrew Conway (2014) of Princeton University, Data Analysis and Statistical Inference by Mine Cetinkaya-Rundel of Duke University, and Passion Driven Statistics by Lisa Dierker of Wesleyan University. While such online courses have existed for quite some time in a variety of fields, the very low-cost structure of many MOOCs has opened the doors to a much larger audience and has increased the accessibility of statistical content.

In 2014, Jeff Leek, Brian Caffo, and I launched the Johns Hopkins Data Science Specialization. This is a sequence of nine courses that intends to provide a "soup-to-nuts" training in data science for people who are highly motivated and have some basic mathematical and computing background. The sequence of the nine courses follows what we believe is the essential "data science process," which is:

1. Formulating a question that can be answered with data
2. Assembling, cleaning, and tidying data relevant to a question
3. Exploring data, checking, and eliminating hypotheses
4. Developing a statistical model
5. Making statistical inference
6. Communicating findings
7. Making the work reproducible

We took these basic steps and designed massive open online format courses around each one of them. We developed this sequence of courses in part to address the growing demand for data science training and education across the globe. Our background as biostatisticians was very closely aligned with the training needs of people interested in data science because, essentially, data science is what we do every single day. Indeed, one curriculum rule that we had was that we couldn't include something if we didn't in fact use it in our own work.

The sequence has a substantial amount of standard statistics content, such as probability and inference, linear models, and machine learning. It also has nonstandard content, such as Git, GitHub, R programming, Shiny, and Markdown.

To date, the sequence has been wildly successful. It averaged 182,507 enrollees per month for the first year in existence. The overall course completion rate was about 6%, and the completion rate among paid enrollees was 67%. In October of 2014, barely 7 months since the start of the specialization, we had 663 learners enroll in the capstone project.

From running the Data Science Specialization, we have learned a number of lessons. Here, I summarize the highlights:

1. *Data science as art and science*. Ironically, although the word "science" appears in the name "data science," there's actually quite a bit about the practice of data science that doesn't really resemble science at all. Much of what statisticians do

in the act of data analysis is intuitive and ad hoc, with each data analysis being viewed as a unique flower.

When attempting to design data analysis assignments that could be graded at scale with tens of thousands of people, we discovered that designing the rubrics for grading these assignments was not trivial. The reason is because our understanding of what makes a "good" analysis different from a bad one is not well articulated in the field of statistics. We could not identify any community-wide understanding of what constitute the components of a good analysis. What are the "correct" methods to use in a given data analysis situation? What is definitely the "wrong" approach? Without such a well-defined framework, useful rubrics are almost impossible to build. We resorted to building fairly minimal assessments, but we believe further research and thinking in this area is sorely needed.

2. *Content* vs. *curation*. Much of the content that we put online is available elsewhere. With YouTube, it is possible to design a course with high-quality videos on almost any topic, and our videos are not really that much better. Furthermore, the subject matter that we were teaching was in no way proprietary. The linear models that we teach are the same linear models taught everywhere else. So what exactly was the value we were providing?

   Effectively, what we provided was a curation of all the knowledge that's out there on the topic of data science (we also added our own quirky spin). Curation is hard, because the curator needs to make definitive choices between what is and is not a core element of a field. But curation is essential for learning a field for the uninitiated.
3. *Skill sets* vs. *certification*. Because we knew that we were not developing a traditional degree program, we knew we had to develop the program in a manner so that the learners could quickly see the value of the program for themselves. This led us to taking a portfolio approach where learners produced things that could be viewed publicly.
4. *New avenues for educational research*. The size and scale of MOOCs created new opportunities for us to conduct research, both by analyzing the data generated by the students enrolled in the courses (and there was a lot of data) and by asking students to volunteer in research studies. For example, we conducted a study of students' abilities to assess the statistical significance of a correlation based on visual displays of data (Fisher, Anderson, Peng, & Leek, 2014). Others have used data generated by MOOCs to study specific pedagogical aspects (e.g., see Guo, Kim, & Rubin, 2014).

13.2.2 Conclusions

As of April 2015, we have had a total of 1158 learners complete the entire specialization, including the capstone project. Given these numbers and our rate of completion for the specialization as a whole, we believe we are on our way to achieving our goal of creating a highly scalable program for training people in data science skills. Of course, this program alone will not be sufficient for all of the data science training needs of society. But we believe that the approach that we've taken, using nonstandard MOOC channels, focusing on skill sets instead of certification, and emphasizing our role in curation, is a rich opportunity for the field of statistics to explore in order to educate the masses about our important work.

13.3 Inference from Big Data: A Cross-Disciplinary Endeavor

Frauke Kreuter
University of Maryland
College Park, MD, USA
University of Mannheim
MannheimGermany
fkreuter@umd.edu

Excitement about big data is visible across many, if not all, disciplines. *Business analytics*, *computational social science*, and *data intensive linguistics* are just a short collection of buzzwords in the behavioral and economic sciences. A similar list could be created for other disciplines as well. Such interest in data is a wonderful opportunity for statistics education, but the burden is on our shoulders to translate the relevant concepts into frameworks useful and applicable in the various disciplines.

Big data have been defined with certain characteristics (e.g., the three v's: veracity, volume, variety) that challenge standard inferential practices. In particular, the lack of a random sample, combined with the large volume of data, often results in a purely descriptive approach. However, standard inference has always faced challenges. For example, beautiful theory justifies drawing inference from a sample to the population, even if the sampling process is complex because it includes multiple stages, as when sampling schools, classrooms within schools, and finally students within classrooms, or because data are collected on the entire clusters at one of the stages.

Unfortunately, the reality of data collection rarely matches the assumptions required of sampling theory. Breakdowns in the data generating process bring into question the well-grounded methods we would like to apply. For the school example above, it is not hard to imagine that principals or teachers might deny access to students or are not able to grant access because of other more important activities going on at school. In household surveys people get sick, are on vacation, or don't want to participate because they lack interest or motivation to answer the survey request. Such breakdowns are not threatening per se. As long as they occur at random, or as long as the mechanism is known and observable, we know how to still create valid inference. Often, neither is the case.

These challenges to standard inferential paradigms have been present for quite some time. However, with the advent of big data, it is tempting to be blindsided by the sheer massiveness of the available data and to overlook the importance of knowing the data generating mechanism here as well. This is not to say that size doesn't solve some of the problems or could help mitigate them. But knowing if that is the case or not requires a deep understanding of the data generating processes if inference is the goal.

Here are some questions that every analysis of big data should ask. These questions should be familiar to all data analysts: What are the proper units of analysis to answer our research question? Are all units we need in the analysis? Are certain units systematically missing? Do some units appear multiple times? Do we have all measures on all units that we need? Whom do the units represent?

Answers to these questions are much easier when the data collection itself is organized by the researcher, ideally with the help of a statistician. If data evolve "organically" and are "found" by the researcher, extra effort will be needed to find the right answers. If Amazon is interested in the correlation between the purchases of two different books on amazon.com, they can of course analyze the population, and no inference problem would be needed. But if, for example, Twitter feeds are examined to measure health or political attitudes, then inference is much harder for some research questions than for others. If trying to answer questions concerning the frequency of health problems or the reoccurrence of health problems, then irregularity of postings, censoring of posts, social desirability, and other issues make the analysis much less straightforward.

Most likely it will be impossible to answer the basic questions posted above without talking to domain experts. Which means in the future, statistics education has to be even more of a cross-disciplinary approach than it currently is. We need to better understand how to teach students to collaborate with researchers in other disciplines, how to better communicate, and how to ask the right questions of big data. When data are collected on humans, the psychologists, sociologists, economists, linguists, etc. will all be able to contribute fundamentally to the understanding of the data generating process. Exciting times head of us!

13.4 Data Science in Secondary Schools

Robert Gould
University of California
Los Angeles, CA, USA
rgould@stat.ucla.edu

Ask a statistician about the future of the profession, and you can expect to hear such trending phrases as "big data," "data science," and maybe even "analytics." These terms are sometimes dismissed as hype—media-friendly buzzwords that repackage what statisticians have always done: find meaning in data.

But to dismiss these as hype misses an important point: data have changed. The open-data movement has brought large, rich, and relevant data to anyone with Internet access (e.g., see data.gov and data.gov.uk). Sensors collect data without human intervention, leading to nonrandom high-density sampling protocols and, often, the creation of "opportunistic" data. Compared to traditional classroom data, these data have great many variables, are rich in high-dimensional relationships, have a complex structure, are daunting in terms of the number of observations, and consist of values that are not always usefully coded as numbers.

Primary and secondary school statistics curricula, in sharp contrast, focus on datasets with one or two variables in which the data have either been randomly sampled from a population (which is often finite though large) or in which units have been randomly assigned to treatment groups. Observational data are included but often pared down to a manageable number of variables and, more often than not, used as a cautionary tale against inferring causality or performing uninterpretable inferential procedures. The Guidelines for Assessment and Instruction in Statistics Education (GAISE), for example, describes three developmental levels, which culminate in formal statistical inference (Franklin et al., 2007). Students learn descriptive statistics and exploratory techniques early on as a means of preparing them for the more complex concepts that support inference.

Why, then, should schools include data that are messy and complex and don't fit into the learning trajectory established by the curriculum? Because these modern data are data that students see every day. For example, students who play online games have data collected about their performance, and these data are shared with other players. Students who wait for a bus might see data displays predicting the arrival of the next bus. Swipe cards, closed-circuit cameras, and large public databases mean that much data about students are stored on the Internet. Whether they want to or not, students are already engaged with modern data. This naive engagement can be dangerous, since modern data raise ethical questions about privacy, confidentiality, and anonymity. But modern data are exciting because they provide opportunities. Many datasets are sufficiently complex that it is realistic to motivate students by reminding them that they can discover something that no one else knows. Further, because of the abundance of technology for sharing, sending, and analyzing data, students who know just a little about analyzing data will be advantaged over those who know nothing.

Working with the Mobilize project, a National Science Foundation-funded partnership between the University of California, Los Angeles Statistics Department; the Graduate School of Education and Information Sciences; and the Los Angeles Unified School District, I′ve had the opportunity to think about how students might interact with modern data. Mobilize provides students with a suite of software and curricular materials to allow them to carry out "participatory sensing campaigns" (Tangmunarunkit et al., 2015). Participatory sensing is a data collection paradigm in which students, acting as human sensors, collect data about their communities and their environment as they go about their everyday life (Burke et al., 2006). For example, students might collect data on where they discard trash and then use data on the location of trash and recycling bins to draw conclusions about how recycling can be improved.

Students can use the *Mobilize dashboard* to visually investigate patterns in the data they have collected. The dashboard allows students to quickly and easily visualize multidimensional relationships between variables. For instance, in one scenario students have seen that classmates who ate cereal in the morning at home tended to rate cereal as a healthier food than did classmates who ate cereal late at night at home. While the dashboard is helpful for discovering patterns, formulating hypotheses, and framing statistical questions, students also learn to use R, via Rstudio, to develop statistical models for these data in a more deliberate and reproducible fashion.

The data collected through participatory sensing are not "big," but they are undeniably modern and share many important characteristics with "big data." For instance, they are collected through a deliberate and yet nonrandom procedure, and they are complexly structured and consist of a variety of types: numerical, categorical, date, time, location, text, and image. These data provide a detailed picture of aspects of students' lives. Interpreting these data requires not only the usual grasp of fundamental statistical notions of distribution and variability but also a sense of the exact constitution of the "inferential universe." Participatory sensing data can therefore serve as a bridge toward formal statistical inference, while in the meantime providing students with interesting and rewarding insights.

When data from a completed campaign are first displayed in the classroom on an interactive map (using the geocoding provided by the smartphone), there is often a collective gasp in the classroom when students see their own daily patterns revealed. (For privacy purposes, and to help make students aware of privacy matters, data are not visible until a student explicitly shares the data, and even then teachers have the ability to purge the data of problematic observations before the data are displayed.) The abstract conversation about privacy and confidentiality is made concrete, and even students whom we may have presumed to be jaded regarding privacy are startled by just how much a collection of observations can reveal when they realize others could deduce where they live and where they spend their time, based on, say, where they are eating their meals and snacks.

These projects underline the need to know more about how students think about modern data, how they can be taught to reason with such data, and what learning trajectories we should design. Kreuter, elsewhere in this chapter, reminds us that thinking carefully about the underlying assumptions behind a statistical analysis will never go out of style, regardless of whether data are big or small. I suspect that all of the statistical concepts and skills taught with "traditional data" will be extremely valuable when learning to analyze modern data. But I also suspect this set of concepts will be too small.

Here is my personal short list of topics that are rarely taught at the secondary level but are potentially accessible and useful: programming, algorithmic thinking, smoothing, nonlinear modeling, goodness of fit, kriging (smoothing and interpolating across spatial processes), classification and regression trees, and density estimation. Each of these is a big idea, but I'd encourage researchers to help us understand just how soon we can introduce these topics and at what level of detail.

13.5 Some Questions About Teaching Computation in Statistics

Randall Pruim
Department of Mathematics and Statistics
Calvin College
Grand Rapids, MI, USA
rpruim@calvin.edu

It is not enough to be convinced that the use of computation by both teachers and students is an important part of statistics education, nor even to be committed to modifying our courses and curricula in keeping with this conviction. We need to do more than merely use computation; we must use it effectively, inspiringly, and in a way that prepares students to learn more than we teach. But how do we do that? And what do educators need to learn in order to do this well?

Below I discuss several questions about teaching statistics. Some of them may be hard questions or may require significant refinement to make them tenable for study. Some answers may be informed by data we do not currently have. But they are important questions, and the answers will go a long way to shaping the evolution of statistics curricula and educational practice over the next decade.

13.5.1 How and When Should Statistics Students Learn the Computational Tools and Skills they Require?

Consider the following from the ASA Undergraduate Curriculum Guidelines (UCG) published by ASA (American Statistical Association Undergraduate Guidelines Workshop, 2014):

> The additional need to think with data in the context of answering a statistical question represents the most salient change since the prior guidelines were endorsed in 2000. Adding these data science topics to the curriculum necessitates developing data, computing, and visualization capacities that complement more traditional mathematically oriented statistical skills.

It now seems that nearly everyone agrees that computational proficiency is an important goal for both undergraduate and graduate statistics programs and that most current programs are deficient in meeting this goal. There is also increasing pressure to introduce computational aspects of statistics in courses for nonmajors. There is much less agreement about what steps should be taken to improve the situation. This is a rather broad question, so let's break it down into some more focused questions.

13.5.2 What Programming Skills Do Statistics Students Need?

According to UCG , thinking with data is a cognitive skill that requires some technical skills in data, computing, and visualization. The challenge is in determining which skills to teach and how to teach them in a way that prepares students to continue to learn new things upon completion of their programs. While obtaining some level of proficiency with particular technological skills is important for completing the task at hand, having a conceptual framework and the confidence necessary to continue to learn new data technologies as required is even more valuable, especially since it will never be possible to teach students all the computational skills that would be beneficial in the context of a statistics program.

Enumerating key concepts required to think with data, and understanding how students come to learn them, is an important prerequisite to understanding the implications of curricular and pedagogical decisions aimed at developing this capacity.

13.5.3 What Can We Learn from the Computer Science Education Community?

It would be naive to think that statisticians have nothing to learn from computer science educators about how to teach computation. It would also be naive to assume that both groups have identical goals for the computational training of students. But there can be little doubt that statistics education researchers interested in how to improve the computational abilities of statistics students should be aware of the work that has been done in the computer science education community and maintain a dialogue with computer science education researchers about current thoughts, work, and trends.

It may require some effort to build bridges between these two communities, but computer scientists are also interested in data science, big data, and many of the other buzzwords of the day, so there are more potential conversation starters than ever before.

13.5.4 Has the "Probability and Mathematical Statistics" Sequence Become Antiquated?

Most of the textbooks in this area still reflect an outline of topics that goes back to classic texts such as Hogg and Craig (1959). Few include significant treatment of newer computationally intensive methods (e.g., randomization testing, bootstrap methods, numerical Bayesian procedures); present methods for handling large, complex, or unsanitized data; or take advantage of computational tools to treat familiar topics in different ways. Most enforce a fairly clean separation between a probability course and a statistics course (which generally assumes a previous probability course).

For more than a decade, there have been calls to rethink which elements of statistical theory are most important for undergraduates to master in a climate where computation both enables and requires statisticians to work differently (e.g., in a session entitled "Is the Math Stat course obsolete?" at the 2003 Joint Statistics Meetings (Rossman & Chance, 2003)), but a decade later, it appears we have not reached a consensus. Some more recent books (e.g., Chihara & Hesterberg, 2011; Pruim, 2011) are moving things in directions that were not available when Hogg and Craig's texts were first published, including computational methods and the theory that supports them, but it would be possible to push things further in this direction.

The growing emphasis on data, computational, and visualization skills is one reason that now is an important time for the mathematical statistics course to be reevaluated with an eye toward determining the most important foundational topics for the next generation of statisticians.

13.5.5 How Well Do We Understand the Statistics Education Landscape?

The Mathematical Association of America recently conducted a 5-year study of calculus instruction funded by the NSF with primary goals to (1) improve our understanding of the demographics of students who enroll in calculus and (2) to measure the impact of the various characteristics of calculus classes that are believed to influence student success.

David Bressoud (former president of the MAA and PI for the study) concluded that the study "revealed that Calculus I, as taught in our colleges and universities, is extremely efficient at lowering student confidence, enjoyment of mathematics, and desire to continue in a field that requires further mathematics (Bressoud & Rasmussen, 2014)." At the same time, the study identified seven characteristics of calculus instruction at institutions that "bucked this trend".

Perhaps it is time to launch a similarly ambitious study of statistics education. Informed by the approach and results of the MAA study, and broadening the scope beyond the United States, a similarly comprehensive investigation of undergraduate statistics instruction could be very informative and provide a much better view of the landscape than we currently have.

13.6 To Bayes or Not to Bayes? (The Answer Is Yes)

Jeff Witmer
Oberlin College
Oberlin, OH, USA
jeff.witmer@oberlin.edu

The undergraduate statistics curriculum is built on the philosophical base of frequentist reasoning. This school of thought is so pervasive that most students are not aware that there is an alternative: Bayesian reasoning.

A Bayesian analysis of a medical trial (to take a specific example) focuses on what most people would consider to be the natural question of interest: "Does the drug work?" In contrast, traditional (i.e., frequentist) inference is developed around an indirect question:

"If this drug is not effective, how likely would it have been to see data such as the data that arose?" No one getting a prescription from a doctor wonders how data might look if the drug doesn't work; instead they wonder what the chances are that the drug will work for them. A Bayesian analyst asks "How likely is it that the effect of the drug is positive, given the data?" and might conclude "Given the data, I believe the probability that the drug works is 97%".

Students in frequentist-based statistics courses often want to interpret a *p*-value as the probability that the null hypothesis is true but are admonished by their professors that this is not correct reasoning. A Bayesian, on the other hand, is allowed to, and indeed must, talk about the probability that the null hypothesis is true (e.g., there is a 3% chance, given the data, that the drug has no effect).

If Bayesian inference provides a direct answer to the question of interest, then why is this not the dominant school of thought in statistics? There are several reasons that frequentism has held center stage for many years. (1) Until recently Bayesians were limited by difficult mathematics and therefore often found themselves restricted to working on a small class of problems. (2) Although some people might have no difficulty imagining a study being repeated an infinite number of times under a given (null hypothesis) scenario, they are unwilling or unable to imagine the effect of a drug as a random quantity. (3) Some people are uncomfortable with the idea that two statisticians can look at the same data and reach somewhat different conclusions just because they started with different prior beliefs; e.g., a second statistician studying the drug experiment might say "Given the data, I believe the probability that the drug works is 98%." (4) There has been little research into how students reason in a Bayesian context, so those who want to teach Bayesian methods might be unsure of what path to take.

Times are changing. Recent advances in computing power and software development have led to big changes in the application of Bayesian statistics. Computers have made possible Markov chain Monte Carlo (MCMC) methods that allow Bayesians to solve problems that were once considered intractable. Today, a Bayesian can fit flexible and complex models that allow one to tackle a wide class of problems.

Regarding (2), that we should treat the parameter of interest (in our case, the effectiveness of a drug) as a fixed, unknown quantity and imagine repeated samples of hypothetical data, one can become quite comfortable with the concept of a parameter as being unknown and thus subject to a probability distribution. Indeed, people assign probabilities to unobserved, but fixed, quantities all the time. For example, toss a coin and let it fall on the floor, but step on it before seeing whether it landed heads up or tails up. Pretty much everyone at this point is willing to say "The probability of heads is ½" despite there being nothing random after the coin is tossed. If I can't see the coin, then its status (as heads up or tails up) is effectively random to me.

Regarding (3), there are many things that can be said. One is that the proper use of prior information should inform statistical and scientific inference. Another is that frequentists are happy to use a directional (versus a typical, nondirectional) alternative hypothesis when prior information tells them that only one direction is plausible. But beyond this is the fact that as more and more evidence accumulates, Bayesians who start with rather different prior beliefs will converge to the same posterior beliefs, which renders moot the objection that scientists should not disagree.

If one is convinced that Bayesian statistics is worth teaching, then how should it be done? Software plays a crucial role in Bayesian practice and thus in Bayesian teaching. In recent years tools such as JAGS and Stan have been introduced that make MCMC somewhat easier to use.

A reality check is in order. MCMC has changed the world of Bayesian methodology, yet as of 2016 we don't have the kind of user-friendly, menu-driven implementations of MCMC that a novice would find easy to use. But this is changing; for example, see the BEST website (Bayesian estimation supersedes the *t*-test, at www.sumsar.net/best_online) to get a taste of how Bayesian methods can be used. Despite a paucity of user-friendly software, some of us teach Bayesian methods with MCMC in undergraduate courses. In the Bayesian course that I teach to undergraduates, I go into a fair amount of detail regarding how MCMC works. However, one can use MCMC without knowing exactly what the computer is doing, much as one can use a *t*-table without first deriving the *t* density. Thus, it is certainly possible to teach a Bayesian course that has exactly the same prerequisites as does a standard introductory frequentist course.

Any university student majoring in statistics should learn about Bayesian methods and MCMC (see also Cobb, this chapter). A nonmajor taking an introductory course (which is often the student's last formal statistics course) should be exposed to Bayes' theorem and the rudiments of Bayesian inference, as one can expect to see more use of Bayesian methods across many fields of study in the years ahead, continuing a trend that began when MCMC became a workable tool.

The fact that Bayesian reasoning is more natural than frequentist reasoning and that computers are making Bayesian methods increasingly accessible leads me to expect continued growth in the teaching of Bayesian reasoning in the years ahead. As ongoing research in statistics leads to better understanding of how students learn, I hope that more attention is paid to Bayesian reasoning. But mostly I hope that educators will teach Bayesian methods to their students.

13.7 Getting to Bayes in Our First Course: Education Research Can Lead the Way

George W. Cobb
Mt. Holyoke College
South Hadley, MA, USA
gcobb@mtholyoke.edu

In 2016 Bayesian inference and Bayesian hierarchical models occupy a major position within applied statistics. However, when it comes to Bayesian thinking, our introductory courses and corresponding research on teaching and learning are still back where things stood 45 years ago when I was in graduate school. I confess that it took me 25 years to overcome my own resistance to Bayes and another 10 to find a way to teach it in a first course, but I'm confident that those who study how we and our students understand data can get to Bayes a lot faster than I did. In what follows, I describe five obstacles, five things that Bayes was *not*: In practice, the Bayesian approach was considered (1) *not objective*, (2) *not computable*, and so (3) *not broadly applicable*. For the elementary course, the Bayesian approach was considered (4) *not accessible* and (5) *not mainstream*.

The next sections use my own conversion to Bayes as a vehicle to discuss (1–3) and then tell how I dealt with (4 and 5) to include Bayesian thinking in a first statistics course, and finally, I conclude with a wish list of five areas where I hope for aggressive research by those who have an interest in statistics and a background in cognitive science.

My conversion to Bayesian thinking. In my first job after college, I worked as a computer programmer and earned a master's degree in the Department of Biometry at the Medical College of Virginia. No one used or taught Bayesian methods. Using a prior distribution made the methods "not objective" and so not scientific. Moreover, the computations required for any but the simplest applications were beyond the capacity of the computers and numerical methods of the day because they involved high-dimensional integrals. In the early 1970s at Harvard, the applied courses were equally Bayes-free. I encountered Bayesian thinking only in a single theory course, and there only as a method for deriving admissible estimators, which were called "Bayes" estimators but were not really Bayesian. My only experience with a genuine use of Bayesian thinking came from reading the Mosteller and Wallace (1964) study of the authorship of the *Federalist Papers*.

A lot changed during the next 20 years. Box and Tiao (1973) got around many computational problems by using conjugate priors and presented a Bayesian approach to many traditional applied problems such as ANOVA and regression. Reanalysis using a range of prior distributions made it possible to assess sensitivity to choice of prior and so moderated concerns about subjectivity. Most dramatically, various versions of Markov chain Monte Carlo (MCMC) methods were adapted to compute posterior distributions for a rapidly growing set of applied problems.

As a result of these and related developments, Bayesian data analysis has become mainstream as one major part of statistical practice. However, teaching Bayesian thinking in a first statistics course is still rare.

My conversion to teaching Bayesian data analysis. For decades, one of the principal objections to teaching Bayes has been that "Bayes is not mainstream." That remains true of introductory courses, but it is certainly no longer true of professional practice. These days Bayesian data analysis (e.g., multilevel models) is definitely mainstream. As I see it, there remains one major obstacle: Bayesian methods are not thought to be accessible at an elementary level. It was only after I found a way around that obstacle that I began teaching Bayes in a first statistics course.

My experience in the classroom convinced me that the essence of Bayesian logic is intuitive and that the main obstacle to teaching Bayes is not conceptual, but technical. In the next section, I suggest six research questions related to those convictions. By way of introduction, I first summarize the approach I have used to introduce Bayesian applications in a first course, (1) compute conditional probabilities by restricting the sample space, and (2) substitute Laplace for Bayes:

1. *Compute conditional probabilities by restricting the sample space*. The traditional approach to Bayesian inference depends on the formal definition of conditional probability and, most importantly, on the denominator in $P(A|B) = P(A \text{ and } B)/P(B)$. My experience tells me that all probabilities are conditional, defined by choice of sample space, and that $P(A|B)$ is the primitive concept, with $P(A \text{ and } B)$ as a corollary, as in the logic of tree diagrams: $P(A \text{ and } B) = P(B)P(A|B)$. (See research topic c below.)
2. *Substitute Laplace for Bayes*. The traditional approach via Bayes' theorem, viz., $P(q|X = x) = P(X = x|q)\, P(q)/P(X = x)$, puts an unnecessary focus on the denominator $P(X = x)$, which is a multiple integral over the entire parameter space. Laplace put it more simply by ignoring the denominator: $P(q|X = x)$ is proportional to $P(X = x|q)$. His principle, which I call "Laplace's data duplication principle," supports all of Bayesian thinking. In my paraphrase of Laplace, "A parameter value is believable to the extent that it can reproduce the observed data value." In the hope of making Laplace's principle concrete and intuitive, I have come to rely on a "Russian roulette" algorithm. The name comes from Kahn (1955) and refers to evaluating conditional probabilities by "killing" outcomes that fail to satisfy the condition. My adaptation to create a Bayesian posterior for parameter q given data y_{obs} is a four-step process (generate, simulate, compare, and estimate):

   (a) *Generate* a random value of the parameter q, according to the prior.
   (b) *Simulate* a data value y_{rand} according to $P(y|q)$.
   (c) *Compare*: Is $y_{rand} = y_{obs}$? If no, kill q. If yes, keep q.
   (d) *Estimate*: Values of q are saved in proportion to how frequently they reproduce the observed value $y_{obs.}$ Thus the saved values estimate the posterior.

This algorithm is horrendously inefficient, but I find that it is an intuitive way to explain Bayesian thinking. (See research topic d below.)

Five topics for research. As I see it, taking Bayes seriously opens the door to myriad unexplored research questions we need to pursue. Here are five:

1. *Use of probability to quantify uncertain knowledge*. To what extent and in what ways is this subjective use of probability intuitive? In what ways does it lead to misconceptions? (Compare interpretations of Bayesian posterior intervals and confidence intervals. Students often misinterpret confidence intervals as Bayesian posterior intervals.)
2. *Predictive* versus *postdictive uses of probability*. This pre-post distinction is due to A.P. Dempster (1964). Pre: Assume a fair coin, which you plan to toss ten times. Before the toss, the *pre*dictive probability of ten heads is 1/1024. Post:

Now assume that all ten tosses came up heads. The probability 1/1024 is now *post*dictive. What does this mean? Fisher used p-values as a measure of surprise, as evidence *against* the initial model. Laplace assumed that $P(X = x|q)$ is proportional to $P(q|X = x)$: the observed x was used as evidence *in favor of q* values. These two interpretations are Bayesian (Laplace) and non-Bayesian (Fisher). Both are important and valid. How can research help us figure out how to teach them both?

3. *Conditional probability: dump the definition*? Which approach to $P(A|B)$ is more intuitive: $P(A)$ with the sample space restricted to B, and with $P(A \text{ and } B)$ as a corollary, or $P(A|B) = P(A \text{ and } B)/P(B)$?
4. *Laplace's* "*data duplication principle*." We need research: In what contexts and to what degree is it intuitive that $P(q|X = x)$ is proportional to $P(X = x|q)$?
5. *Incorporating prior information and sensitivity analysis*. If we accept Bayesian logic as an important approach to data analysis, we need to address the challenge of subjectivity, by changing the prior and tracking the consequences. Is this intuitive?

Over a lifetime as a statistician, I have become convinced that Bayesian logic will become more and more important as one essential part of thinking about data. In the past, research in statistics education has tended to track what we have already been teaching, sometimes in support, sometimes in question, but so far, not as a radical challenge. In suggesting a challenge, I am unsure about how best to integrate Bayesian thinking into our current teaching. The right research can help us decide.

References

American Statistical Association Undergraduate Guidelines Workgroup. (2014). *2014 curriculum guidelines for undergraduate programs in statistical science*. Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/education/curriculumguidelines.cfm.

Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. New York: Wiley Classics.

Bressoud, D., & Rasmussen, C. (2014). Seven characteristics of successful calculus programs. *Notices of the American Mathematical Society, 62*(2), 144–146.

Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., et al. (2006). *Participatory sensing*. UCLA: Center for Embedded Network Sensing. Retrieved from http://escholarship.org/uc/item/19h777qd.

Chihara, L., & Hesterberg, T. (2011). *Mathematical statistics with resampling and R*. Wiley.

Conway, D. (2014). Data science through the lens of social science [Video Lecture]. *20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), New York 2014*. Retrieved from http://videolectures.net/kdd2014_conway_social_science/.

Dempster, A. P. (1964). On the difficulties inherent in Fisher's fiducial argument. *Journal of the American Statistical Association, 59*(305), 56–66.

Fisher, A., Anderson, G. B., Peng, R. D., & Leek, J. T. (2014). A randomized trial in a massive online open course shows people don't know what a statistically significant relationship looks like, but they can learn. *PerrJ, 2*, e589.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A preK-12 curriculum framework*. Alexandria, VA: American Statistical Association.

Guo, P., Kim, J., & Rubin, R. (2014). *How video production affects student engagement: An empirical study of MOOC videos. In Proceedings of the 1st ACM Conference on Learning @ Scale*. New York: ACM Digital Library. http://dl.acm.org/citation.cfm?id=2566239.

Hogg, R. V., & Craig, A. T. (1959). *Introduction to mathematical statistics*. New York: The Macmillan Company.

Kahn, H. (1955). *Use of different Monte Carlo sampling techniques*. Santa Monica: The Rand Corporation.

Lund, S., Manyika, J., Nyquist, S., Mendonca, L., & Ramaswamy, S. (2013, July). *Game changers: Five opportunities for US growth and renewal*. McKinsey Global Institute.

Mosteller, F., & Wallace, D. L. (1964). *Inference and disputed authorship: The federalist*. Reading, MA: Addison Wesley.

Pierson, S. (2014). *Largest U.S. Master's programs in statistics and biostatistics*. ASA Community blog, posted Feb 09, 2014. community.amstat.org/blogs/.

Pruim, R. (2011). *Foundations and applications of statistics*. American Mathematical Society.

Rossman, A., & Chance, B. (2003). Notes from the Joint Statistics Meetings session entitled "Is the Math Stat course obsolete?" Retrieved from http://www.amstat.org/sections/educ/MathStatObsolete.pdf.

Tangmunarunkit, H., Hsieh, C. K., Longstaff, B., Nolen, S., Jenkins, J., Ketcham, C., et al. (2015). Ohmage: A general and extensible end-to-end participatory sensing platform. *ACM Transactions on Intelligent Systems and Technology, 6*(3), Article 38.

Chapter 14
Building Capacity in Statistics Teacher Education

João Pedro da Ponte and Jennifer Noll

Abstract This chapter focuses on building capacity in statistics teacher education, with a twofold purpose. First, this chapter reviews prior research into the professional development of teachers of statistics (elementary, middle, secondary, and tertiary teachers). This provides an overview of different approaches to professional development for teachers of statistics, challenges statistics educators face when building professional development programs, and lessons learned that can inform directions for future research. Second, this chapter looks forward by outlining major challenges, gaps in our knowledge base, and important directions for future research in the professional development of teachers of statistics. We outline a vision for the future of professional development that focuses on building capacity through a strongly integrated focus on practice, content, and technology.

Keywords Building capacity • Professional development • Statistics teacher education

14.1 Introduction

The primary goal of this chapter is to look forward toward building capacity in statistics teacher education. By building capacity, we mean a conceptual approach to the professional development of prospective and practicing teachers and teacher educators, which address both quality and quantity with the goal of improving statistics education. For this, it is necessary to take a broad view of "statistics teachers." Many of those who teach statistics are mathematics teachers, and some are science,

J.P. da Ponte (✉)
Instituto de Educação, Universidade de Lisboa, Lisbon, Portugal
e-mail: jpponte@ie.ulisboa.pt

J. Noll
Fairborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR, USA
e-mail: noll@pdx.edu

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_14

or computer science teachers who teach statistics within the context of their discipline. In primary schools, many teachers are generalists, teaching statistics as part of the mathematics curriculum. We therefore include in our discussion any primary, middle, secondary, and tertiary teachers who teach statistics.

Although teacher development may lead to positive changes in teaching practice (Franklin, 2014), there is a past record of inefficacy of conventional statistics teacher education (Ponte, 2011). The professional development of teachers of statistics is therefore a very prominent issue, presenting many challenges. However, before we begin the process of looking forward, we first look at the existing research on professional development for teachers of statistics. What has been the typical environment for this professional development? In what ways have both in-service and preservice teachers been targeted? How has professional development focused on teachers' content knowledge, practice, or the integration of the two? What has been the role of technology in professional development? What does prior professional development research tell us about the issues our community needs to address? What are the challenges educators face when developing and implementing professional development?

This chapter has six sections. The first section is a synthesis of the different environments of professional development of statistics teachers described in the literature. The second focuses on teachers' statistical content knowledge. This has been discussed in Chap. 10 of this handbook, but here we highlight how research that focuses on teachers' statistical content knowledge may inform changes in professional development. In the third section, the role of technology in professional development is addressed, and in the fourth the focus is on teachers' practice. The fifth section addresses the intersection of content, practice, and technology in professional development. Finally, we leave the reader with lessons learned through a synthesis of the big ideas and challenges that emerge in the research literature and identify directions for future research toward building capacity in statistics teacher education.

14.2 Environment of Professional Development

A review of the literature reveals many different approaches to professional development. It may be long term or short term; it may include mandated coursework for preservice or in-service teachers or optional courses or workshops for both pre- and in-service teachers developed through national organizations and grant initiatives. Professional development may be offered in face-to-face, online, or hybrid formats. It may focus on statistical content, on the practice of teaching statistics, or both. This section synthesizes key features of current professional development work.

The majority of the research into professional development reviewed for this chapter focused on mathematics teachers who also teach statistics, including elementary, middle, and secondary school teachers. Most focused on the professional development of preservice teachers during their mandated university coursework

(Batanero, Gea, Díaz, & Cañadas, 2014; Batanero, Godino, & Roa, 2004; Browning, Goss, & Smith, 2014; Canada, 2006; Confrey, Makar, & Kazak, 2004; Dolor & Noll, 2015; Garfield & Everson, 2009; Groth & Xu, 2011; Heaton & Mickelson, 2002; Leavy, 2010; Lee & Hollebrands, 2008a, 2008b; Lee & Nickell, 2014). Sometimes this focus on university courses also included in-service teachers taking coursework for their continuing education requirements (Confrey et al., 2004; Groth & Xu, 2011; Madden, 2014; McClain, 2008; Meletiou-Mavrotheris & Mavrotheris, 2007; Meletiou-Mavrotheris, Paparistodemou, & Stylianou, 2009; Serradó-Bayés, Meletiou-Mavrotheris, & Paparistodemou, 2014).

We hypothesize two reasons as to why more of research has focused within the context of preservice teachers. First, preservice teachers are the most readily available group for researchers working in university settings. Second, it is plausible that more researchers and educators are concerned that preservice teachers can be as prepared as possible before they enter their classrooms. Most of the professional development work occurred as part of university courses (for pre- and/or in-service teachers), typically over a 10- to 15-week period with approximately 3 h per week of meeting time. However, there were also a few short-term professional development studies (e.g., 4 days in Madden, 2011) or longer-term studies (e.g., 6 months in Henriques & Ponte, 2014; 1 year in Wassong & Biehler, 2014; and 2 years in Makar, 2010).

A few professional development approaches used online or hybrid formats (Garfield & Everson, 2009; Meletiou-Mavrotheris & Mavrotheris, 2007; Meletiou-Mavrotheris et al., 2009; Serradó-Bayés et al., 2014). For example, Meletiou-Mavrotheris and her colleagues had success in collaborations with teachers in Spain, Cyprus, Greece, Norway, and Ireland. They developed *EarlyStatistics*, an online professional development program. Participating teachers from different European countries formed a virtual community of practice (Lave & Wenger, 1991), developing their own projects based on a wide array of resources, colleagues, and discussions. They found that an international collaboration of teachers, through the creation of online communities, is able to share challenges many teachers face as well as new approaches to these challenges. These collaborations can be helpful to teachers in a range of contexts.

There are additional advantages of online and hybrid formats. For instance, online platforms have the potential to reach in-service teachers who live in rural locations and cannot regularly attend university classes. In addition, teachers have increasingly busy workdays and may be involved in after-school activities, so online professional development that is more accommodating to their busy schedules may encourage more participation. However, there are still challenges to this format. For example, Meletiou-Mavrotheris et al. (2009) and Serradó-Bayés et al. (2014) noted that misunderstanding of statistical ideas or strategies occurred as a result of cultural or linguistic barriers or stark differences in national or state curricula. In addition, challenges can exist with online formats if participation in group online chat is not maintained, as mentioned in Meletiou-Mavrotheris and Serradó-Bayés (2012). Elements in the design of online courses relate to how new material is introduced, the kinds of experiences it provides to learners, the nature and amount of interaction between facilitators and learners and among learners themselves, and the content and structure of the course. The work described here suggests that building capacity in statistics teacher education through using technology to reach more teachers is an important component.

At the university level, the current state of preparation of teachers of statistics looks quite different. In contrast to the preparation of teachers in primary, middle, and secondary schools, there is no mandated preparation for those who go on to teach in higher education. As a result, there is very little research investigating the professional development for teachers of undergraduate statistics courses, who are often graduate teaching assistants. The few professional development approaches described in the literature are from institutions that have created courses to prepare their graduate teaching assistants for teaching introductory statistics (see Froelich, Duckworth, & Stephenson, 2005; Gelman, 2005; Harkness & Rosenberg, 2005). These professional development courses were one term (15 weeks in duration) meeting for only 1 h per week. They had two primary purposes: to provide graduate teaching assistants with examples of successful and unsuccessful teaching or how to act professionally in the classroom (not necessarily specific to the teaching of statistics) and to provide activities and curriculum materials for teaching introductory statistics, with examples of how to implement activities in the classroom. For example, Froelich et al. (2005) described an apprenticeship model where graduate students begin their apprenticeship as graders and lab recitation leaders. These graduate teaching assistants worked closely with the course instructor to ensure consistency and to learn more about course material. Those who did well with these beginning teaching activities became course instructors. The department supported these teaching assistants by providing homework, answer keys, a syllabus, and lab assignments. Teaching assistants also received a file of old exams and quizzes, lecture notes, and power point presentations to model their coursework.

Thompson and Johnson (2010) argue that easing into teaching statistics through this type of apprenticeship model (graders to tutors to lab recitation leaders to teachers) is important for better preparing our future undergraduate statistics teachers. However, while these statistics-specific workshops for graduate teaching assistants should help prepare them for their teaching responsibilities and make them more comfortable in the classroom, they are not necessarily informing them about current research in how undergraduates learn statistics. These courses help graduate students to more easily replicate the current teaching of statistics by their faculty members. Further research in this area could provide the community with empirical evidence about graduate teaching assistants' practice and subsequent impact on student learning.

14.3 Approaches to Building Capacity through Developing Teachers' Content Knowledge

A major focus of this chapter is on building capacity in teachers' statistical content knowledge via professional development. Overwhelmingly, the research suggests that professional development courses are the most effective when they focus on creating the sort of learning environment that statistics educators would like to see teachers develop in their classrooms (Canada, 2006; Lee & Hollebrands, 2011; Makar & Confrey, 2005; Meletiou-Mavrotheris et al., 2009; Pfannkuch, 2008; Pratt, Davies, & Connor, 2011). Suggestions for the design of teacher education courses include:

1. Build on teachers' experiences and prior knowledge (Henriques & Ponte, 2014; Makar & Confrey, 2005).
2. Have teachers create lesson plans and questions for students (Lee & Nickell, 2014; Pfannkuch & Ben-Zvi, 2011).
3. Have teachers examine and respond to student work (Confrey et al., 2004; Lee & Nickell, 2014; Makar & Confrey, 2005; Makar & Fielding-Wells, 2011).
4. Incorporate the big ideas of statistics into coursework (Browning et al., 2014; Madden, 2011; Makar & Confrey, 2005; McClain, 2008; Pfannkuch, 2008; Pfannkuch & Ben-Zvi, 2011).

A major strand of research concerning teachers' professional development in statistics puts the notion of *statistical investigation* at the very center of the classroom activity (incorporating the fourth suggestion mentioned above). In particular, the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) reports (pre-K-12 report: Franklin et al., 2005; College report: American Statistical Association, 2016) have been influential documents, used by researchers from many countries as a way of structuring statistics activities in professional development settings (Browning et al., 2014; Dolor & Noll, 2015; Garfield & Everson, 2009; Green & Blankenship, 2013; Heaton & Mickelson, 2002; Henriques & Ponte, 2014; Lee & Hollebrands, 2008b; Meletiou-Mavrotheris & Mavrotheris, 2007; McClain, 2008). The GAISE reports emphasize the process of statistical investigations suggesting that teaching statistical ideas begin with the general approach to (1) formulate a research question, (2) collect data that could help answer the question, (3) analyze the data collected, and (4) interpret the results. The pre-K-12 GAISE report also contains three developmental levels that outline a progression in statistical literacy and highlight the notion of variability.

Statistical investigations are seen as being at the heart of the activity of statisticians (see Chaps. 1 and 4 in this volume) and essential to provide students with a real sense of the activity. Statistical investigations are also a suitable way to introduce students to the specific concepts, representations, processes, and procedures that constitute the statistics curriculum, from primary school to university. It has been argued that prospective teachers must learn about statistical investigations in order to use them later in their practice (Heaton & Mickelson, 2002; Makar &

Fielding-Wells, 2011), to support children's reasoning with data (Santos & Ponte, 2014). They also need a disposition to teach in a way that recognizes the substantive knowledge in the different issues and topics addressed through investigations (Heaton & Mickelson, 2002).

Research has reported some successes with professional development that focuses on developing content knowledge through statistical investigations. For example, Leavy (2006) noted positive changes in prospective elementary teachers' awareness of the importance of data explorations, increased attention to ideas of distribution, and increased ability to use alternative measures and representations when comparing data. However, a number of challenges are also reported in professional development settings where statistical investigations were the focus. Heaton and Mickelson (2002) noted that prospective elementary teachers focused mainly on technical components of investigation, losing site of the overall statistical process. Santos and Ponte (2014) found the elementary teacher in their case study appeared focused on a sequence of techniques to apply rather than the overall statistical process. Leavy (2006) also noted that many prospective elementary teachers focused on summary features of data and ignored variation.

The pre-K-12 GAISE report (Franklin et al., 2005) highlights the importance of the whole notion of what a statistical investigation is, beginning with the formulation of a research question followed by the research design and data collection. However, some researchers have attempted to streamline parts of the statistical investigation process in professional development programs to more efficiently build teacher content expertise. In particular, some researchers have discussed simplifying the data collection process by using data already available on the internet. For example, Hall (2011) suggested using primary and secondary data to simplify the problem formulation and data collection phases. In addition, Batanero et al. (2014) used UNESCO data in their workshop for prospective high school teachers.

Makar (2010) conducted a long-term professional development project in teaching statistical investigations with in-service elementary teachers. Her work involved 23 primary teachers over a 2-year period. While teachers did indicate positive perspectives about what inquiry could mean for their students, they expressed many concerns such as curriculum/time pressures, classroom management issues, managing open-ended problems, and content knowledge of the statistical process.

Confrey et al. (2004) had another approach to using statistical investigations for professional development. In their work, they examined preservice middle and high school teacher's content knowledge in a professional development course that focused on statistical investigations about equity and high-stakes testing. Teachers were asked to examine data on their students' scores on high-stakes tests and interpret the results in relation to educational equity. There were improvements of prospective teachers' posttest scores suggesting the teachers in their study improved their understanding of how to investigate and apply statistics in an authentic context. Their work also addressed three of the four points recommended above for professional development of teachers—building from prior experiences, responding to student work, and focusing on the big ideas of statistics.

Metz (2010) also focused on statistical investigations, applying the pre-K-12 GAISE report guidelines (Franklin et al., 2005) as well as of the College Board of Mathematical Sciences (CBMS) (Blair, Kirkman, & Maxwell, 2013) for a professional development course for prospective primary and middle school teachers. However, Metz's work was different from other statistical investigation approaches. Metz wanted prospective teachers to make connections between content in the investigations and what they might teach. Metz asked teachers to map statistical activities to the GAISE developmental levels (A, B, and C) and consider how an activity could be modified or revised for use at a lower or higher level. This approach differs from other work focused on statistical investigations in that there was a more explicit attempt to connect investigations to teaching practice and to understand a particular statistical investigation through the three levels of GAISE. This work has the promise of helping prospective teachers think about where their students may be coming from (an earlier GAISE level of statistical literacy) and where they may be going to (an advanced GAISE level). However, Metz focused on teachers' evaluations of the course, which were generally positive, but provided little in the way of informing the community about whether teachers improved their content knowledge or their knowledge of implementing GAISE ideas at the different developmental levels.

A different approach to building capacity in teachers' content knowledge during professional development was the incorporation of the Realistic Mathematics Education (RME) and guided reinvention framework used by Dolor and Noll (2015). They incorporated the theory of RME that mathematics should be learned naturally as students work to solve problems that are realistic to them (see Gravemeijer, 2004 and Chap. 16 in this volume) and apply this approach to work with statistics teachers. Pre- and in-service high school teachers enrolled in a 10-week professional development course focused on statistical investigation and the big ideas of statistics. Classroom activities were designed around an instructional sequence focused on reinventing significant ideas of hypothesis testing. The goal was not for them to develop hypothesis testing as might be seen in an introductory statistics text but rather to support and motivate the teachers in (re)inventing an informal hypothesis test for categorical data. Several of the teachers in this study successfully reinvented an informal test similar to a chi-square test.

The professional development of teachers of undergraduate statistics courses at colleges and universities deserves special mention. In the United States, undergraduate statistics courses are on the rise at community colleges, colleges, and universities (see CBMS, 2013 and Chap. 2 in this volume), and this is likely to be true at universities across the world as more courses require basic statistical reasoning and as statistical literacy skills become increasingly important in a technological, data-driven, global economy. Those teaching these courses include graduate teaching assistants teaching statistics for the first time as well as part-time and full-time faculty. They may have a profound impact on undergraduate students' learning in statistics. Moore (2005) suggested that in the United States, future statisticians may only become interested in statistics because they are required to take an introductory statistics course during their undergraduate career. If they have a good experience in their first course, they may, as a result, change their studies to statistics. Moore argued that if these students enjoy their experience and get hooked on data analysis, they might choose to focus on a career in statistics or statistics education. This is unlikely to happen if these undergraduate students do not have knowledgeable and experienced teachers of statistics. Yet there is little research investigating statistics teachers at tertiary level. Like the majority of secondary mathematics teachers, many graduate students complete their bachelor's degree in mathematics and may never have taken a statistics course before. They may be entering graduate statistics programs and taking their first statistics course while at the same time beginning their first teaching assignments in statistics. They may lack either the content knowledge or the knowledge of teaching (or both) to be well prepared to teach these undergraduate courses (see Moore, 2005 and Chap. 10 in this volume). In one of the few studies investigating graduate teaching assistants, Noll (2011) found that many did not have foundational knowledge of sampling distributions. Given that graduate teaching assistants become our future community college and university faculty, it is imperative that more research is done to study their content knowledge and create professional development to develop that content knowledge.

In summary, much of the research in professional development reviewed for this chapter focused on statistical investigations, in some form or another, as an approach for building capacity in teachers' statistical content knowledge. However, this approach has significant challenges in its transposition to classroom practice. The research revealed that many teachers focus on procedural features of investigations or get stuck in problem formulation and data collection phases. While teachers do tend to improve their content knowledge from these classes, it does not appear that the impact is significant enough for them to be able to model statistical investigations in a robust way in their classrooms. Many teachers do not have an understanding of underlying big ideas of statistics (Garfield & Ben-Zvi, 2008) or a strong sense of the statistical process. As a consequence, the central or marginal role that statistical investigations may have in classrooms is still an open question. Statistical investigations have a great promise but also present recognized challenges. To value the integrity of the statistics investigation process is an important aim but perhaps one that is not possible to fulfill in many cases given the external constraints.

Even when the conditions are favorable, there remain questions as to whether investigations are the most efficient way to teach all statistics processes and ideas. For example, statistical investigations take a significant time to implement in teacher education programs (as well as in classrooms), and therefore the time that can be devoted to developing a deep and connected understanding of the ideas of formulating important research questions, designing appropriate data collection methods, learning robust data analysis approaches, and devoting time to the important work of interpreting results is limited. For example, in Dolor and Noll's (2015) study on reinventing a big statistical idea of hypothesis testing, considerable time was spent to get teachers to the point of reinventing one type of hypothesis test, and there was no evidence that the teachers could then translate these ideas to other hypothesis tests or have a deeper understanding of the process of statistical inference. At some point the research community needs to address the issue of limited time in professional development, and whether or not lessons learned from research in statistical investigations in professional development can be developed more quickly and deeply or scaled up to address the multitude of statistical ideas teachers need to better understand.

14.4 Approaches to Building Capacity through Developing Teachers' Content and Technological Knowledge

Today the practice of statistics is inconceivable without the use of technology. Apart from making the simplest graphs and representations, all work with authentic statistics presupposes the use of technology (Ben-Zvi, 2000; Biehler, Ben-Zvi, Bakker, & Makar, 2013; Chance, Ben-Zvi, Garfield, & Medina, 2007). Statisticians, statistics educators (e.g., Gould, 2010; Lee & Hollebrands, 2008a, 2008b; Madden, 2011), and national organizations (e.g., National Council of Teachers of Mathematics (NCTM), 2000; American Statistical Association (ASA), 2016; Franklin et al., 2005) argue that technology can and should be used in the teaching and learning of statistics for middle, secondary, and university students. However, many teachers of statistics are likely to have limited experience of working with educational technologies designed for the statistics classroom. Even those teachers who have experience using statistical software packages have probably used technology in rudimentary ways, such as to do computations more quickly, rather than using technology in more dynamic ways, such as learning how to construct different visual representations of data through the dragging and dropping of attributes on the screen.

Dynamic technologies such as TinkerPlots (Konold & Miller, 2011) or Fathom (Finzer, 2012) are not software packages that would typically be found in middle, secondary, and tertiary schools; yet these technologies have been recommended as important to support students' development of key statistical ideas (e.g., Biehler et al., 2013; Lee & Hollebrands, 2008a, 2008b; Lee & Nickell, 2014; Lee et al., 2014; Madden, 2011). We have already articulated how research suggests that teachers are likely to teach what they know and what they are comfortable with (e.g., Lee & Hollebrands, 2011, 2008a, 2008b; Pratt et al., 2011). Thus, in order to build capacity in statistics teacher education, we need to understand how professional development that integrates technologies might have an impact on what teachers do in the classroom. This section investigates these issues.

Much of the professional development research has focused on teachers' statistical content knowledge. However, there is a smaller body of research studying the impact of integrating technology into professional development courses. Madden (2011) suggested that TinkerPlots (Konold & Miller, 2011) could be used for comparing variability in distributions of data to create "technologically provocative tasks" for teachers, that is, tasks that create epistemological obstacles or intellectual conflicts for teachers to support their development of new statistical ideas. Madden

(2014) provided examples of tasks that appeared to support both pre- and in-service secondary teachers' development of technological pedagogical statistical knowledge (TPSK, see Chap. 10 in this volume). Madden focused on an intentional sequencing of the environment, moving from the physical environment to the computer environment as well as technologically provocative tasks, to create an environment where teachers experience cognitive conflict but at the same time feel comfortable to take intellectual risks.

In another study, focused on the big statistical idea of variability, Browning et al. (2014) investigated prospective elementary teachers in a course that used TinkerPlots and strove to follow the GAISE report recommendations (Franklin et al., 2005). The authors suggested that the technology provided prospective teachers with a "conceptual way of appropriately attending to measures of variability" (p. 1). They conjectured that this would translate into them being better able to support students' development of concepts of variability in the classroom and to use technology to support that development.

McClain (2008) developed an instructional sequence to tie together middle school statistics curricula more coherently under the premise that the integration of computer tools is a critical component. The middle school teachers had computer tools that allowed them to manipulate, order, partition, and organize data. Tasks began with assessing teachers' understanding of different types of graphs and how to make arguments from graphical displays. As the sequence progressed, the tasks demanded more nuance from the participants. Several of the tasks required comparing two distributions of data. The intention was to create distributions that would problematize the comparison of means. Teacher talk shifted from a focus on the procedures of measures of central tendency to conversations grounded in the data and a consideration of context. McClain argued that such interventions may improve middle school teachers' statistical understanding of significant ideas as well as building an instructional agenda supporting the development of significant statistics ideas in their students.

Meletiou-Mavrotheris et al. (2009) investigated ways to support elementary school in-service teachers in Cyprus to implement technology in their classrooms. They identified factors that hindered teachers from implementing technology: lack of specific instructions for integrating technology into the teaching process and curriculum, lack of recommendations for what software to use, and a shortage of computers or suitable software. The authors designed a professional development curriculum that integrated TinkerPlots with core curricular ideas and data-centered activities using contexts familiar to children. Their work was geared toward supporting teachers' opportunities to model and investigate these data-centered activities with technology with the hope that this would promote their ability to enact similar work within their classrooms. The activities included tasks for teachers to solve with and without TinkerPlots. For example, teachers were given a data chart on life expectancies for women and men in several European countries. They were asked to compare life expectancies first without technology and then with TinkerPlots. They observed that without technology, the teachers primarily focused on numeric strategies for solving problems. When using TinkerPlots on the same problem, they went beyond the bounds of what they were asked to do and considered multiple lenses for viewing the data. This finding that using technology helped support the teachers to view a problem in new and more robust ways supports the findings of McClain (2008). Meletiou-Mavrotheris et al. (2009) also observed that the teachers personally liked using the technology because they suggested it would give them a chance to bring current data into their classroom (e.g., importing internet data into TinkerPlots) rather than using outdated data available in textbooks.

Lee and Hollebrands (2008b) state that "whether technology will enhance or hinder student learning depends on teachers' decisions when using technology tools that are often based on knowledge gained during teacher preparation programs" (p. 326). Their research focused on the use of simulation, dynamic technologies, and data analysis in an integrated way. The materials were designed to support teachers' development of big statistical ideas and understanding of how statistics and technology can serve as a tool for answering questions connected to data where deterministic statements cannot be made. Similarly, Lee and Nickell's (2014) research focused on middle and secondary preservice teachers and was driven by the following research questions: "How do teachers use dynamic software tools to engage in statistical ideas and how may their work be preparing them for pedagogical decisions when teaching students statistics in their future classroom?" (p. 1). A significant finding from their research is that after such a course, focused on using dynamic technologies to teach statistics, the teachers may still not be ready to use dynamic aspects of the technology in their teaching. They noted that many teachers used technology to automate computations and the creation of graphs but tended to produce graphs they were already familiar with and did not always experiment with the rich variety of data representations that were available to them with the technology. This also made it challenging for them to learn new things about their data.

Lee et al. (2014) studied prospective and in-service elementary, middle, and secondary school teachers and found a similar result. They observed that many of the teachers in their study appeared to have routinized their work and suggested that this "may indicate that the development of their statistical knowledge and technological statistical knowledge is not moving beyond viewing the tool as a way to automate the creation of graphs and computations and engaged in basic types of transnumeration activities" (p. 19). These findings stand in contrast to those described earlier (Madden, 2014; McClain, 2008; Meletiou-Mavrotheris et al., 2009).

Wassong and Biehler (2014) implemented a 1-year professional development course in Germany for experienced secondary mathematics teachers who are also responsible for mentoring other teachers in their country. The first 4 months of the course focused on having teachers investigate big ideas of statistics (as other professional development focused on content has done) while integrating technology into those investigations. Wassong and Biehler found that the teachers had very little prior experiences with technology. The software appeared to help their learning of statistical concepts as well as how to explain statistical ideas. Podworny and Biehler (2014) also did research in a professional development course focused on prospective secondary teachers taking a basic statistics course for their degree program. They used a statistical investigations approach with TinkerPlots. Their work focused on the topic of simulations and hypothesis testing. They found that about half of the preservice teachers preferred to use the simulation scheme during their work in TinkerPlots as a support for retrospectively documenting what they did during their simulation. However, most of the preservice teachers did not want to use a simulation scheme for planning out their hypothesis testing approach.

Pratt et al. (2011) suggested three major challenges relating to the integration of technology: (1) teachers' lack of comfort with, or knowledge of, the technology; (2) priority of technology at teacher, school, and assessment levels; and (3) teachers poorly implementing technology in the classroom. Much of the professional development research summarized in this section addressed the first challenge. It provides examples about how to potentially increase teacher knowledge on the use of digital technologies for doing statistics. Some studies (Browning et al., 2014; Madden, 2011, 2014; McClain, 2008; Meletiou-Mavrotheris et al., 2009) noted positive outcomes in teachers' abilities to learn new statistical ideas via the technology; however, a couple studies (Lee & Nickell, 2014; Lee et al., 2014) noted that after concluding the professional development courses, teachers still tended to use the technology in routinized ways (ones they were used to and comfortable with) rather than alternative approaches that might be better suited to the data at hand. Thus, the statistics education community still needs to address how to improve teachers' knowledge of dynamic statistical software packages. In addition, Wassong and Biehler (2014) questioned how many technologies can we expect teachers to learn and at what level of depth. This is a pertinent question that the statistics education community needs to address especially given the lack of time in most professional development courses.

The other two challenges raised by Pratt et al. (2011), priority of technology and teachers poorly implementing technology, have received little attention from the research. There needs to be further research in the professional development of teachers of statistics that focuses on the priority placed on technology by teachers, schools, and assessment materials as well as research that focuses on how teachers' actually implement technology into the statistics curriculum. Wassong and Biehler (2014) suggested one issue related to poor implementation: when symbolic representations in digital formats differ from those presented in textbooks, teachers may struggle with their own content knowledge as well as how to effectively work around those issues within their classrooms. Poor implementation of technology is not a focus of research in part because most focuses on preservice teachers' content knowledge and misses an opportunity to integrate issues of content (as well as content learned with technology) with the practice of teaching. We need more research that integrates all three components—content, technology, and practice—so that teachers have experience bringing statistical investigations using software to fruition in their own classrooms. Effective implementation of technology in the schools depends on teachers' beliefs about the role of technology in statistics. The statistics education community needs to build studies that focus on ways to build teacher beliefs and teacher practices, spending more time examining the work teachers do with statistical software packages in the classroom. In particular, design studies (e.g., Kelly, Lesh, & Baek, 2008) that examine teacher practice before and after professional development are badly needed. Limited resources at schools may also limit the ability to implement technology into statistics courses. The statistics education community will need to address ways to make important technologies for investigating statistics more broadly available.

14.5 Approaches to Building Capacity through Simultaneously Developing Teachers' Practice and Content Knowledge

This section focuses on professional development research that emphasized the integration of practice and content knowledge. Professional development focused on content knowledge has tended to have a unifying theme of statistical investigation or big ideas of statistics. In our review, professional development projects focused on practice also had a common theme of engaging teachers in inquiry in their own classrooms, placing the teacher as researcher (or action research, Mertler, 2016). Researchers have suggested that professional development is more effective when it occurs within the context of teachers work in the classroom (Meletiou-Mavrotheris & Serradó-Bayés, 2012; Ponte, 2011).

Lesson studies were a key feature of several of the professional development studies reviewed and are one way of asking teachers to engage in structured inquiry into their classroom lessons (Garfield, delMas, & Chance, 2005, 2007; Roback, Chance, Legler, & Moore, 2006). A lesson study broadly involves three major steps (Murata, 2011): (1) identifying a topic of common interest to the teachers and planning of a lesson around topic, (2) the observation of the lesson taught by one of the teachers of the group, and (3) reflection and follow-up. Lesson plans attempt to consider not only the curriculum and resources available but also the current research on student learning of the topic. Observations focus on how students solve the tasks proposed. Reflections include all teachers analyzing what happened in class, focusing on what students are doing and experiencing. Analysis may lead to the reformulation of the lesson, changing the strategy, the materials, the tasks, or the questions posed to the students. The teaching cycle may be repeated several times (Murata, 2011).

The work of Garfield et al. (2005, 2007) was conducted at the university level where the teachers were two experienced faculty members, a teacher, and four graduate teaching assistants. In a research by Roback et al. (2006), lesson study was used in an advanced statistics course addressing the concept of sampling distribution in the context of goodness-of-fit tests. Both groups of researchers reported that lesson study was very useful for better understanding student thinking and learning, promoting better collaboration among students (both in large and small group settings), better collaboration among teachers, and improved lessons because of the benefits of multiple perspectives.

Leavy (2010) used lesson study as part of a professional development course for prospective primary teachers. Lesson study was used as a way for the teachers to learn content while informing practice via the planning of lessons and the implementation of those lessons in their classrooms with their students. Leavy (2010) concluded that the participants demonstrated proficiency in informal inferential reasoning (content knowledge) but had difficulties developing pedagogical contexts to promote such reasoning in primary school students. Overall, lesson study was regarded as a useful form of classroom-based inquiry and focusing on lessons as the unit of analysis advanced the participants' development of statistical and pedagogical knowledge.

A benefit to lesson study is that it provides teachers with a collaborative experience with an intensive focus on teaching a particular topic. Participants have a unique opportunity to observe lessons and collaborate with other teachers of statistics as well as having teacher educator mentors to reflect on the outcomes of a lesson. While this type of collaborative effort by teachers is common in Japan, it is unusual to find this type of collaboration in many other countries. A major challenge of lesson study is that it is very time-consuming both as an approach to teaching and to teacher education (Garfield et al., 2005, 2007).

An approach with similar features to lesson study is that of Teacher Professional Development Cycle (TPDC) developed by Souza, Lopes, and Pfannkuch (2015). This model built on Ponte's (2011) work and was developed out of a concern to create professional development that improves teachers' specialized content knowledge, pedagogical content knowledge, and professional knowledge. Souza et al. (2015) created a five-phase model: orientation, exploration, application, analysis, and reflection. Like lesson study, the orientation phase focuses on teachers identifying problems in their practice that they wish to pursue. In the exploration phase, teachers investigate relevant research and plan a lesson around the identified problems. Teachers also plan how they will collect data about their lessons for the reflection phase. In the analysis and reflection stage, teachers use formative assessment and conduct a self-assessment and an analysis of their students' thinking. As with lesson study, this work is collaborative in that teachers work together through the cycles supported by a facilitator who acts as a partner. The researchers implemented their approach with middle school teachers and reported case studies on two teachers. They suggest that these teachers improved their statistics knowledge and changed their approach to teaching statistics. In addition, they developed their ability to plan a lesson and to work collaboratively with other teachers.

Ponte's (2012) national professional development program for elementary and secondary in-service teachers was rooted in practice to help prepare teachers to teach using a new curriculum for basic education introduced in Portugal in 2007. An important theme in the new curriculum was data handling. The professional development program was organized in 25 h of face-to-face sessions and an additional 25 h of independent work. There were five main program themes: (1) orientation toward practice through familiarizing teachers with the new curriculum, (2) focus on students' learning through the creation and implementation of lesson plans aligned with the goals of the new curriculum, (3) collaboration, (4) practitioner research through collection of action research data in teachers' classrooms, and (5) change of professional culture through collaboration, discussions, and reflections about the lessons they taught. Overall, this professional development was situated in teachers' own practice because they carried out a lesson in their classes, collected data from their classrooms, and analyzed this data. The hope was that the focus would support teachers' capacity to teach statistics with the new curriculum, and the resulting work showed some success. Teachers were proud of the activities they created, and there was much support for the new curriculum.

There are also a few studies that have been attempting to integrate practice and content knowledge in both preservice and in-service statistics teacher education through video case analysis or case analysis (Groth & Xu, 2011; Lee & Hollebrands, 2008a, 2008b). For example, Groth and Xu (2011) illustrated two situations, one referring to a case discussion among a group of prospective secondary mathematics teachers in the United States and the other concerning a teaching research activity involving a group of in-service junior high school teachers in China. In case analyses, educators use a teacher's (hypothetical or real) lesson plan, class activity, student work, and teacher reflections as the basis for professional development work (and, when available, video of classroom activity or students working on problems). Groth and Xu (2011) indicated that the discussions that ensue in this type of professional development environment can help improve teachers' content knowledge as they consider the statistics in the class activities, the examples of students' work, and the teachers' reflections/notes. This type of professional development environment can also improve teachers' content-specific pedagogical knowledge and general pedagogical knowledge as they reflect on and discuss aspects of students' work and the teacher's decisions in the classroom.

14.6 Approaches to Building Capacity through Developing Teachers' Practice, Content Knowledge, and Technological Knowledge

A relatively small number of studies have attempted to integrate teachers' practice, content knowledge, and technological content knowledge into professional development courses. Henriques and Ponte (2014) suggested that in addition to a focus on specialized content knowledge for teaching statistics, such professional development must also integrate other recommendations made for introductory statistics classrooms, such as using exploratory data analysis (EDA) and new digital technologies. They suggest that, beyond robust statistical content knowledge, in-depth understanding of how students develop statistical ideas and how to use technology tools in productive ways that support students' statistical development, statistics teachers need to hold productive beliefs and conceptions about statistics.

Their research studied the implementation of professional development grounded in a social theory of learning that focuses on building on the experiences of participating teachers over a period of 6 months. The teacher educators focused on three primary goals: (1) building an awareness of the importance of statistical reasoning, (2) presenting recommendations and guidelines for what to include in introductory statistics courses, and (3) ways to develop ideas suggested in guidelines and opportunities to plan and conduct lessons using TinkerPlots and to select instructional activities that support student learning. In this way, the professional development focused on teacher knowledge development related to activities that build knowledge through an examination of teacher practice. Henriques and Ponte (2014) suggest that this type of professional development is an effective way to support teachers, giving them a "better appreciation of what teaching statistics is" and helping teachers align their classrooms to current curriculum standards (p. 5).

Meletiou-Mavrotheris and her colleagues developed a professional development course that also integrated practice and content with technology (Meletiou-Mavrotheris, Mavrotheris, & Paparistodemou, 2011; Meletiou-Mavrotheris & Serradó-Bayés, 2012). The course was organized around the statistical problem-solving process and sought to involve teachers in reflecting on and discussing the different stages of this process and the related specialized content knowledge. According to the authors, the *EarlyStatistics* course "aims at helping teachers improve their pedagogical and content knowledge of statistics through exposure to web-based educational tools and resources, and cross-cultural exchange of experiences and ideas" (p. 3). The course was developed in 13 weeks and included seven modules. The first 6–7 weeks focused on statistical investigations using TinkerPlots. In the second 6–7 weeks, the professional development focused on classroom implementation issues, as teachers adapt and develop the material provided and use it in their practice with the support of the facilitators. This professional development also uses video case analysis to discuss issues of content and practice. The course concluded with teachers reporting their experiences and collective discussion and reflection.

Lee and Hollebrands (2008a, 2008b)) integrated practice, content, and technology through video case analysis of students in a classroom working on statistical investigations using TinkerPlots. Teachers in the professional development had to focus on student work with the aim that teacher discussions around student work using TinkerPlots would support their development of statistical content, statistical pedagogical content, and statistical technological pedagogical content knowledge.

Another approach that connects practice, content, and technology is that of Kuzle and Biehler (2015) that aimed to prepare 12 experienced mathematics teachers to become teacher educators ("mentor teachers"). These teachers participated in a 5-month continuous professional development course in which a major feature was the development and implementation of a short professional development activity lasting for one afternoon. The aim of this course was to further the participants' professional knowledge of teaching statistics using digital tools and to develop their knowledge and competences for designing and implementing their own professional development courses in statistics. The course included three components: (1) knowledge for teaching statistics, (2) knowledge for designing effective professional development activities, and (3) knowledge about models of teachers' professional knowledge for teaching statistics with digital tools.

The professional development research described in this section focused on practice, content knowledge, and technological content knowledge and appeared to show much success in supporting teachers to improve their practice in the classroom. Practice, content knowledge, and technological content knowledge are key areas for building capacity in statistics teacher education. Thus, professional development that can implement all three key aspects is particularly salient since we know that teachers are likely to teach what they have experienced.

14.7 Looking Ahead

The professional development of teachers of statistics is incredibly an important work with which the statistics education community needs to engage. Teachers' content knowledge, as well as what they do in classrooms, impacts on student learning. In this chapter, we have synthesized how the statistics education community has been attending to issues in professional development. In particular, our review of the literature revealed that more of the research has focused on preservice teachers' content knowledge (see Chap. 10 in this volume). The strengths we found in these studies include grounding research in national guidelines such as GAISE reports, focusing on statistical investigations or big ideas of statistics, and utilizing dynamic educational software such as Fathom or TinkerPlots. Some of the challenges these studies revealed include the need for more time than typical university coursework provides to make substantial improvements (which go beyond improving procedural knowledge) in content knowledge and knowledge of statistical software packages. Future research needs to focus on how to more efficiently improve statistical content knowledge and knowledge of statistical software packages.

Somewhat less prominent, but still quite prevalent in the literature, was research focused on in-service teachers' practice and content knowledge. Much of this research was also carried out through university courses where teachers began working through statistical investigations, focusing on content knowledge, and then later the course beginning a focus on how to bring statistical investigations into their own classrooms. Teachers worked either independently or in groups with other teachers to develop lesson plans, try lessons out in their classrooms, reflect on their lessons, discuss outcomes with peers and educators, and sometimes go through another round of lessons after revisions to the original lesson were made. These professional development courses tended to focus on content and practice by integrating statistical content knowledge gained at the beginning of the course with practice in which teachers act as researchers by implementing statistical investigations into their own classrooms. Rather than approach the work of teachers from a theoretical perspective within the classroom, the emphasis appears to be moving toward grounding professional development in the practice of teaching, that is, using activities that teachers are likely to use in their classrooms, stimulating discussion around student artifacts, focusing on student centered routines for classroom activities, and, in some cases, situating professional development in the practice of teaching through classroom-based action research.

A few studies, most notably Henriques and Ponte (2014), Lee and Hollebrands (2008a, 2008b), Meletiou-Mavrotheris et al. (2011), and Meletiou-Mavrotheris and Serradó-Bayés (2012), integrated statistical investigations, statistical software technology, and practice. The use of video analysis by Lee and Hollebrands (2008a, 2008b) represents a novel approach in our field for bringing aspects of practice into preservice professional development. We need to see more of this work. In general, any work that brings the focus on student thinking during statistical investigations and has teachers examining student work has the potential to both develop teachers' content knowledge and help prepare teachers for the work of responding to student work. There needs to be more research into the professional development of teachers of statistics that integrates all three aspects (content, technology, and practice). If we expect teachers to competently teach statistical investigations, the big ideas of statistics, and how to use statistical software appropriately, then we must engage them in this work in their professional development. The same is true with respect to practice: teachers cannot be expected to successfully implement robust statistical investigations using technology into their classrooms without disciplined efforts to focus on their practice and particularly on students' activity in the classroom. Including practice into professional development is key if we want to develop teachers who will critically examine and reflect upon the impact of their lessons on their students' learning. In addition, statistics educators need to study further the impact of national and state curriculums on the work teachers do in the classroom and try to find ways to generate conversation that bring statistical investigation, technology, and practice in mandated curricula to classrooms.

Given the challenges that lie ahead in building capacity in statistics teacher education, we also need to consider alternative approaches that were not prevalent or in absent from the research we reviewed. We argue that professional development for preservice teachers cannot just be restricted to content or content integrated with technology if we are to best prepare our future teachers of statistics. We also suggest that statistics teacher educators build collaborations with schools so that their preservice teachers have more opportunities: spending time in actual classrooms observing students' working on statistical investigations, reflecting on student work in the classroom, and having introductory experiences creating and implementing lessons prior to their first assignments. Co-teaching or mentoring may be ways to bring practice into the education of future teachers of statistics, for example, research that explores co-teaching where a preservice and in-service teacher work together, as in the innovative approach to lesson study described by Cajkler and Wood (2016a, 2016b). While this type of mentoring occurs in student teaching activities, it does not appear to be well researched and does not always appear to be integrated with a particular content course. We need content courses where preservice teachers act as an apprentice and in-service teachers as mentors. The experience may in fact be a mutually supportive relationship if pre- and in-service teachers have an opportunity to explore statistical investigations together and then to explore ways to bring those investigations into their classrooms. The act of co-preparing lessons, enacting lessons, reflecting on class sessions, and grading and assessing student work provide opportunities to discuss content, students' statistical reasoning, and practice.

While the quality of undergraduate instruction has not always been a priority for colleges and universities, particularly research institutions, we are now at a point in time when colleges and universities are being held more accountable for the quality of undergraduate education. Statistics educators need to do more to engage in collaborations with statisticians to build momentum for improving teaching and learning of introductory statistics at the college level. We need to study graduate students, part-time faculty, and professors to better understand issues related to content knowledge and practice in these groups. At community colleges in the United States, we find instructors with master's degrees in mathematics teaching introductory statistics having never taken any statistics as part of their coursework. Thus, the statistics education community must work to make inroads with these varied communities and instructors with varied backgrounds. Models of professional development for primary through secondary school contain aspects that could be applied to professional development of teachers of statistics in higher education. Professional development using lesson study, co-teaching, mentoring, or video case studies are models that could be used with teachers of statistics in higher education. There is currently little research investigating the practice of teachers of statistics in higher education. Some educators have informally shared what they do to prepare graduate students to teach introductory statistics. For example, Garfield and Everson (2009) described their professional development for teachers of statistics using the GAISE report as a guiding framework. This course served not only graduate teaching assistants but also those who will go on to teach high school statistics. Their work focuses on specific issues in the teaching of introductory statistics and is geared to build future instructors' content knowledge as well as specific pedagogical knowledge, through understanding students' misconceptions and research-based study of students' statistical development. However, faculty at community colleges and universities are completely missing from the professional development literature. Without collaborations with statisticians, this group is likely to continue to remain missing because there is no mandated professional development. Statisticians and statistics educators need to develop guidelines for the practice of teaching undergraduate statistics courses.

References

American Statistical Association. (2016). *Guidelines for assessment and instruction in statistics education (GAISE): College report*. Alexandria, VA: Author.

Batanero, C., Gea, M. M., Díaz, C., & Cañadas, G. R. (2014). Building high school prospective teachers' knowledge to teach correlation and regression. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference* *on Teaching Statistics, ICOTS9, July, 2014, Flagstaff, AZ, USA*. Voorburg, The Netherlands: International Statistical Institute.

Batanero, C., Godino, J. D., & Roa, R. (2004). Training teachers to teach probability. *Journal of Statistics Education, 12*(1).

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1&2), 127–155.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–690). New York: Springer.

Blair, R., Kirkman, E. E., & Maxwell, J. W. (2013). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States (Fall 2010 conference board of mathematical sciences survey)*. American Mathematical Society.

Browning, C., Goss, J., & Smith, D. (2014). Statistical knowledge for teaching: Elementary preservice teachers. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July 2014, Flagstaff, AZ, USA*. International Statistical Institute: Voorburg, The Netherlands.

Cajkler, W., & Wood, P. (2016a). Lesson study and pedagogic literacy in initial teacher education: Challenging reductive models. *British Journal of Educational Studies, 64*(4), 503–521.

Cajkler, W., & Wood, P. (2016b). Adapting 'lesson study' to investigate classroom pedagogy in initial teacher education: What student-teachers think. *Cambridge Journal of Education, 46*(1), 1–18.

Canada, D. (2006). Elementary prospective teachers' conceptions of variation in a probability context. *Statistics Education Research Journal, 5*(1), 36–64.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1), Article 2.

Confrey, J., Makar, K., & Kazak, S. (2004). Undertaking data analysis of student outcomes as professional development. *International Journal of Mathematics Education, 36*(1), 32–40.

Dolor, J., & Noll, J. (2015). Using guided reinvention to develop teachers' understanding of hypothesis testing concepts. *Statistics Education Research Journal, 14*(1), 60–89.

Finzer, W. (2012). Fathom Dynamic Data Software (Version 2.13) [Computer software]. Emeryville, CA: Key Curriculum Press and KCP Technologies.

Franklin, C. (2014). The statistical education of teachers (SET): An American Statistical Association policy document. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July, 2014, Flagstaff, AZ, USA*. International Statistical Institute: Voorburg, The Netherlands.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A preK-12 curriculum framework*. Alexandria, VA: American Statistical Association.

Froelich, A. G., Duckworth, W. M., & Stephenson, W. R. (2005). Training statistics teachers at Iowa State University. *The American Statistician, 59*(12), 8–10.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.

Garfield, J., delMas, R., & Chance, B. (2005). *The impact of Japanese lesson study on teachers of statistics*. Paper presented at the Joint Statistical Meetings, Minneapolis, MN.

Garfield, J., delMas, R., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 117–147). New York, NY: Lawrence Erlbaum.

Garfield, J., & Everson, M. (2009). Preparing teachers of statistics: A graduate course for future teachers. *Journal of Statistics Education, 17*(2).

Gelman, A. (2005). A course on teaching statistics at the university level. *The American Statistician, 59*(1), 4–7.

Gould, R. (2010). Statistics and the modern students. *International Statistical Review, 78*(2), 297–315.
Gravemeijer, K. (2004). *Creating opportunities for students to reinvent mathematics. Proceedings of the 10th International Congress in Mathematics Education* (pp. 4–11). Denmark. Online: http://www.staff.science.uu.nl/~savel101/edsci10/literature/gravemeijer1994.pdf.
Groth, R. E., & Xu, S. (2011). Preparing teachers through case analyses. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education (a Joint ICMI/IASE Study)* (pp. 371–381). New York, NY: Springer.
Green, J. L., & Blankenship, E. E. (2013). Primarily statistics: Developing an introductory statistics course for pre-service elementary teachers. *Journal of Statistics Education, 21*(3).
Hall, J. (2011). Engaging teachers and students with real data: Benefits and challenges. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education (a Joint ICMI/IASE Study)* (pp. 335–346). New York, NY: Springer.
Harkness, W., & Rosenberg, J. (2005). Training graduate students at Penn State University to teach statistics. *The American Statistician, 59*(12), 11–14.
Heaton, R. M., & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education, 5*(1), 35–59.
Henriques, A., & Ponte, J. P. (2014). Preparing teachers to teach statistics: Developing professional knowledge and practice. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July, 2014, Flagstaff, AZ, USA*. International Statistical Institute: Voorburg, The Netherlands.
Kelly, A. E., Lesh, R. A., & Baek, J. Y. (Eds.). (2008). *Handbook of design research methods in mathematics education: Innovations in science, technology, engineering, and mathematics learning and teaching*. New York, NY: Routledge.
Konold, C., & Miller, C. D. (2011). TinkerPlots: Dynamic data exploration (Version 2.2) [Computer software]. Emeryville, CA: Key Curriculum Press.
Kuzle, A., & Biehler, R. (2015). Examining mathematics mentor teachers' practices in professional development courses on teaching data analysis: Implications for mentor teachers' programs. *ZDM Mathematics Education, 47*(1), 39–51.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
Leavy, A. (2006). Using data comparison to support a focus on distribution: Examining preservice teachers' understandings of distribution when engaged in statistical inquiry. *Statistics Education Research Journal, 5*(2), 89–114.
Leavy, A. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning1. *Statistics Education Research Journal, 9*(1), 46–67.
Lee, H. S., & Hollebrands, K. F. (2008a). Preparing to teach data analysis and probability with technology. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference* (pp. 359–369).
Lee, H. S., & Hollebrands, K. F. (2008b). Preparing to teach mathematics with technology: An integrated approach to developing technological pedagogical content knowledge. *Contemporary Issues in Technology and Teacher Education, 8*(4), 326–341.
Lee, H. S., & Hollebrands, K. F. (2011). Characterising and developing teachers' knowledge for teaching statistics with technology. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education. A joint ICMI/IASE Study* (pp. 359–369). Dordrecht, The Netherlands: Springer.
Lee, H. S., Kersaint, G., Harper, S. R., Driskell, S. O., Jones, D. L., Leatham, K. R., et al. (2014). Teachers' use of transnumerations in solving statistical tasks with dynamic statistical software. *Statistics Education Research Journal, 13*(1), 25–52.

Lee, H. S., & Nickell, J. (2014). How a curriculum may develop technological statistical knowledge: A case of teachers examining relationships among variables using Fathom. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July 2014, Flagstaff, AZ, USA*. International Statistical Institute: Voorburg, The Netherlands.

Madden, S. R. (2011). Statistically, technologically, and contextually provocative tasks: Supporting teachers' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1–2), 109–131.

Madden, S. R. (2014). Designing technology-rich learning environments for secondary teachers to explore and prepare to teach statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July, 2014, Flagstaff, AZ, USA*. International Statistical Institute: Voorburg, The Netherlands.

Makar, K. (2010). Teaching primary teachers to teach statistical inquiry: The uniqueness of initial experiences. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics (ICOTS8, July, 2010, Ljubljana, Slovenia)*. Voorburg, The Netherlands: International Statistical Institute.

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27–54.

Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education (a Joint ICMI/IASE Study)* (pp. 347–358). New York, NY: Springer.

McClain, K. (2008). The evolution of teachers' understandings of distribution. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching statistics in school mathematics: Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*.

Meletiou-Mavrotheris, M., & Mavrotheris, E. (2007). Online communities of practice enhancing statistics instruction: The European Project EarlyStatistics. *Electronic Journal of e-Learning, 5*(2), 113–122.

Meletiou-Mavrotheris, M., Mavrotheris, E., & Paparistodemou, E. (2011). Distance learning for teacher professional development in statistics education. *Teaching Statistics, 33*(1), 2–8.

Meletiou-Mavrotheris, M., Paparistodemou, E., & Stylianou, D. (2009). Enhancing statistics instruction in elementary schools: Integrating technology in professional development. *The Montana Mathematics Enthusiast, 6*(1&2), 57–78.

Meletiou-Mavrotheris, M., & Serradó-Bayés, A. (2012). Distance training of mathematics teachers: The EarlyStatistics experience. *RUSC - Revista de Universidad y Sociedad del Conocimiento, 9*(1), 340–353.

Mertler, C. A. (2016). *Action research: Improving schools and empowering educators* (5th ed.). Thousand Oaks, CA: SAGE Publications.

Metz, M. L. (2010). Using GAISE and NCTM standards as frameworks for teaching probability and statistics to prospective elementary and middle school mathematics teachers. *Journal of Statistics Education, 18*(3).

Moore, D. (2005). Preparing graduate students to teach statistics: Introduction. *The American Statistician, 59*(12), 1–3.

Murata, A. (2011). Introduction: Conceptual overview of lesson study. In L. C. Hart, A. Alston, & A. Murata (Eds.), *Lesson study research and practice in mathematics education: Learning together* (pp. 1–12). New York, NY: Springer.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Noll, J. (2011). Graduate teaching assistants' statistical content knowledge of sampling. *Statistics Education Research Journal, 10*(2), 48–74.

Pfannkuch, M. (2008). Training teachers to develop statistical thinking. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*.

Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education (a Joint ICMI/IASE Study)* (pp. 323–333). New York, NY: Springer.

Podworny, S., & Biehler, R. (2014). A learning trajectory on hypothesis testing with TinkerPlots – Design and exploratory evaluation. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July 2014, Flagstaff, AZ, USA*. The Netherlands, International Statistical Institute: Voorburg.

Ponte, J. P. (2011). Preparing teachers to meet the challenges of statistics education. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education (A Joint ICMI/IASE Study)*. New York, NY: Springer.

Ponte, J. P. (2012). A practice-oriented professional development programme to support the introduction of a new mathematics curriculum in Portugal. *Journal of Mathematics Teacher Education, 15*(4), 317–327.

Pratt, D., Davies, N., & Connor, D. (2011). The role of technology in teaching and learning statistics. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education. A Joint ICMI/IASE Study* (pp. 97–107). Dordrecht, The Netherlands: Springer.

Roback, P., Chance, B., Legler, J., & Moore, T. (2006). Applying Japanese lesson study principles to an upper-level undergraduate statistics course. *Journal of Statistics Education, 14*(2).

Santos, R., & Ponte, J. P. (2014). Learning and teaching statistical investigations: A case study of a prospective teacher. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July, 2014, Flagstaff, AZ, USA*. International Statistical Institute: Voorburg, The Netherlands.

Serradó-Bayés, A., Meletiou-Mavrotheris, M., & Paparistodemou, E. (2014). EarlyStatistics: A course for developing teachers' statistics technological and pedagogical content. *Statistique et Enseignement, 5*(1), 5–29.

Souza, L., Lopes, C. E., & Pfannkuch, M. (2015). Collaborative professional development for statistics teaching: A case study of two middle school teachers. *Statistics Education Research Journal, 14*(1), 112–134.

Thompson, H. M., & Johnson, H. L. (2010). How young statistics academics learn to teach statistics. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute.

Wassong, T., & Biehler, R. (2014). The use of technology in a mentor teacher course in statistics education. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July, 2014, Flagstaff, AZ, USA*. International Statistical Institute: Voorburg, The Netherlands.

Chapter 15
Revolutions in Teaching and Learning Statistics: A Collection of Reflections

Robert Gould, Christopher J. Wild, James Baglin, Amelia McNamara, Jim Ridgway, and Kevin McConway

Abstract In the second set of reflective writings (the first set appears in Chap. 13), another group of prominent statisticians and statistics educators considers the impact of recent and future developments on both the statistics curriculum and the ways in which it is taught and learned. The two connecting themes in this group of writings are the ubiquitous use of technology and the uses of data in decision making. All of these writings acknowledge, to a greater or lesser extent, the differing future needs of two groups. As citizens, today's students need to be educated to be critical consumers of data but do not need detailed knowledge of statistical techniques. The much smaller group who will go on to be professionally engaged in the production and analysis of data also need the ability to engage with more technical details. A challenge for statistics education, particularly at school level, is to provide a learning environment which is appropriate for both of these needs, since we do not know the future trajectories of our students.

R. Gould (✉)
Department of Statistics, University of California, Los Angeles, CA, USA
e-mail: rgould@stat.ucla.edu

C.J. Wild
Department of Statistics, The University of Auckland, Auckland, New Zealand
e-mail: c.wild@auckland.ac.nz

J. Baglin
RMIT University, Melbourne, VIC, Australia
e-mail: james.baglin@rmit.edu.au

A. McNamara
Smith College, Northampton, MA, USA
e-mail: amelia.a.mcnamara@gmail.com

J. Ridgway
School of Education, Durham University, Durham, UK
e-mail: jim.ridgway@durham.ac.uk

K. McConway
School of Mathematics and Statistics, The Open University, Milton Keynes, UK
e-mail: kevin.mcconway@open.ac.uk

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_15

Keywords Technology • Big data • Data science • Open data • Journalism • Visualization • Reproducible analysis • Computation • Data analysis

15.1 Introduction

In the second set of reflective writings (the first set appears in Chap. 13), another group of statisticians and statistics educators considers the impact of recent and future developments on both the statistics curriculum and the ways in which it is taught and learned. The two connecting themes in this group of writings are the ubiquitous use of technology and the uses of data in decision making.

Chris Wild (Sect. 15.2) envisages developments in technology which allow all students access to "broad vistas" of how data can be used to investigate real-world problems and support curiosity. James Baglin (Sect. 15.3) focuses attention on the need for students to learn technological skills as consumers and producers of data, while Amelia McNamara (Sect. 15.4) discusses the current "gap" between accessible technologies designed to support the learning of statistics and the more powerful ones used by statisticians to engage in analysis. Jim Ridgway (Sect. 15.5) also highlights the need to educate citizens in the critical use of data but takes a different perspective describing a project in which relatively novice school students explore issues relevant to their own lives using data in the public domain. Kevin McConway (Sect. 15.6) provides a further perspective on the uses of data in the social and political domain by considering the statistics education of journalists.

All of these writings acknowledge, to a greater or lesser extent, the differing future needs of two groups. As citizens, today's students need to be educated to be critical consumers of data but do not need detailed knowledge of statistical techniques. The much smaller group who will go on to be professionally engaged in the production and analysis of data also need the ability to engage with more technical details. A challenge for statistics education, particularly at school level, is to provide a learning environment which is appropriate for both of these needs, since we do not know the future trajectories of our students. The final chapter of this section (Chap. 16), which follows this group of writings, discusses pedagogic approaches to meeting this challenge.

15.2 Lucid Dreams About the Future

Christopher J. Wild
University of Auckland
AucklandNew Zealand
c.wild@auckland.ac.nz

I have a dream of a multitude of students spellbound by the broad vistas of the data landscape. I have a dream of their flying on magic carpets that enable them to swoop effortlessly over this landscape exploring its nooks and crannies in search of its hidden treasures. I have a dream of students empowered to look at data and explore analysis systems and educational environments designed so that, like Alice in Wonderland, they keep crying "*Curiouser and curiouser!*" and have the ability and confidence to go where that curiosity leads. I have a dream of educational and analysis environments designed to leverage the power of "I wonder …?" to draw students into learning more and more—the power of "I wonder why …?," the power of "I wonder what happens if …?," the power of "I wonder what that does?," and the power of "I wonder what's around the next bend or just over the horizon?" I have a dream of software that finesses away the mundane, the mind-numbing, and the soul-destroying difficulties. I have a dream of software that creates rich, virtual data-generating environments that mimic real real-world environments well enough to enable realistic experiential learning in accelerated time frames. I have a dream. But this is not drug-induced fantasy. We are already building systems that are converging on this Utopian dreamscape.

Complementing the dreams are realizations. We are all well aware that the accelerating onslaught of technology is having profound effects on our everyday and workaday lives. But in statistics education, the most profound effect of technology is its affect on what is actually worth learning and by whom. Ours is a fast-changing world of ever-expanding possibilities, where the limits of what machines can do expand inexorably, leaving us ever freer to concentrate on the thinking that is necessarily human. Increasingly, everything that is purely procedural in statistics will be automated in software (a process that will be accelerated by the large numbers of computer scientists in "data science," their every sinew and fiber commanding, "Automate!"). Investigators and data analysts won't need to know the details any more than they now know the details of what is happening under the hoods of their cars. Teaching procedural details, i.e., having students learn to operate particular algorithms (or particular software menus), is teaching short-term, death-dated skills leaving them with little of long-term value. We need to start operating at much higher and more conceptual levels because only the big picture concepts, the fundamental principles and questions, are of real enduring value.

We should be educating large numbers of students (I would argue virtually all students) to think with data and have some facility with conducting and critiquing real-world investigations and much smaller numbers of people who can develop new methodologies and turn them into new tools. For the larger group, we need to get them fast to broad vistas and the big issues to create a sense of possibility and potential for their lives—to open eyes, quicken hearts, liberate the imagination, and empower—then temper this with proper caution. All of this should be facilitated by the best tools available. If struggling to get the right stuff into and out of software chews up an appreciable proportion of these students' time, then it's the wrong software.

To moderate some of the statements above, "short term" is not entirely a negative. Short-term skills can be crucial for that imminent dissertation or that first job.

Additionally, I am not claiming that learning to operate an algorithm (follow instructions) is not a useful skill. It is a very useful skill, but not for operating the algorithms you have been trained on, but for operating new algorithms in the time before they become shrink-wrapped in software. The same applies to learning how to write computer programs (e.g., for data wrangling). Today's complex programming task is tomorrow's mouse-click. These skills deliver their value when they have been mastered to a level where they allow you to do new things that others have not already catered for—bridging the gap between the available and a desired "something more."

I think the key to "experience a lot quickly" is software and educational experiences developed in tandem. The software is designed to make possible capabilities that are educationally desirable. The educational experiences exploit the doors opened by the software. Comprehensive systems are much more desirable than sets of one-off applets. While the latter can be fine for illustrating particular points, they do not foster the forming of broad understandings facilitated by unifying frameworks. Additionally, every time someone reaches for a new system, there is the time sink of figuring out how the new system "thinks." Besides wasting precious time, this can trigger a "too hard" response that inhibits getting started.

We need to distinguish between software that accelerates the speed with which learners can discover new landscapes and is good for occasional users to dip in and out of and software that professionals will immerse themselves in. The former will tend to prioritize areas where the inputs and outputs can be easily understood and not worry about comprehensive coverage, whereas the latter must provide almost everything a professional could want. This inevitably adds to the complexity in human-software interactions and steeper learning curves. Software to accelerate breadth of data-world awareness and empower occasional users, on the other hand, should minimize how many names you have to know before you can get value out of the software and should maximize how far default settings can take you (memories of the what and how of doing things dim surprisingly quickly when not constantly refreshed by regular use).

I have experimented with these ideas in building the iNZight system for data visualization and analysis (www.stat.auckland.ac.nz/~wild/iNZight/) and VIT (Visual Inference Tools, www.stat.auckland.ac.nz/~wild/VIT/), a visualization system for developing inferential concepts. iNZight is heavily used in New Zealand schools, particularly the last 3 years of high school, but has capabilities covering much of undergraduate statistics. The free online MOOC "Data to Insight" (www.stat.auckland.ac.nz/~wild/d2i/4StatEducators/) was produced to prototype a getting-further-faster introductory statistics course leveraging the capabilities for acceleration provided by iNZight and VIT. At one level, iNZight and VIT are prototypes to be learned from when building better getting-further-faster software for the future. But at another, they are also very good systems for teaching and data exploration right here, right now. Although they do not quite usher in the opening dreamscape, they come close enough to show that there is huge potential here and that this is something that is almost within our grasp.

15.2.1 How Can and Should Research Relate to All of This?

The data world is expanding fast but the extent of what we convey has changed very little. If our breadth of view does not start to keep pace with the expansion of the data world, our educational offerings will illuminate an ever-shrinking segment of reality converging to irrelevance, and we will (deservedly) fade into oblivion. To chart our way into that future, we need seers and dreamers, we need entrepreneurs and innovators, we need architects and builders, and we badly need research and researchers.

We need our researchers to take on problems that are bigger and more fundamental but messier and less well understood. Even just crystallizing key research questions constitutes making research contributions of fundamental value. I conclude with a list of some research challenges.

15.2.2 Future-Facing Areas in Need of Substantial Research

- Where is statistics now? What new fields are opening up?
- Of the potential new areas, which are most worth pursuing and why? What is essential versus what is peripheral and why? What don't we need to teach anymore and why?
- Who are the clients for this educational offering (including future employers)? What do they need? What skillsets are required for that? Where in the program should we address that?
- What can machines do and what of statistical thinking is essentially human?
- What difficulties can we circumvent with new technology?

15.3 Teaching Statistics Technology

James Baglin
RMIT University
Melbourne, VIC, Australia
james.baglin@rmit.edu.au

Technology is used in all aspects of preparing, collecting, manipulating, analyzing, summarizing, and visualizing today's large, complex, "big data." It has enabled the growth and development of many computationally intensive methods including bootstrapping, permutation methods, ensemble methods, and Markov chain Monte Carlo methods for Bayesian methods, to name a few (see Chap. 13 for more on Bayesian methods). Gould (2010) was correct in concluding that statistics and technology have become inseparable. This raises an important question for statistics educators: "How do we teach statistics technology?"

Technology for doing statistics fits within the broader domain of statistical computing. Important topics in this area include, but are not limited to, knowing how to use general and statistical programming languages, access databases, manipulate data, use visualization tools, scrape unstructured data from the web, and conduct simulations (Nolan & Temple Lang, 2010a, 2010b). These topics are most relevant to statistics students and are becoming increasingly important to those studying statistics as part of their disciplines. However, at the very least, an introductory statistics student should be familiar with using a common statistical package such as SPSS, SAS, or R. The technology required by the modern statistics student is diverse, and this will likely evolve continually in years to come. The key will not be to teach a specific software or programming language but to develop the ability of students to become lifelong learners of statistics technology.

How to foster the development of this capability in our students remains a pressing challenge. Limited research has explored the development of technological skills in statistics—or so-called technology for "doing" statistics (Baglin & Da Costa, 2013a). The research that does exist looks closely at comparing strategies used to develop statistical package skills in introductory courses (Baglin & Da Costa, 2012, 2013b, 2014). Collectively, this research shows that conceptual understanding and experience using the technology are key ingredients for success. However, statistical package skills are just one example of technology used in modern statistical practice. This must expand to technologies used at all stages of the data investigation process including planning, collecting data, and communicating findings (e.g., visualizations). With this in mind, it was hopeful to have found many examples in the literature of statistics educators creating innovative approaches to better integrate technological skill development into their courses.

In a paper by Hardin et al. (2014), the authors report numerous strategies for teaching data science topics within the statistics curriculum, namely, the incorporation of technology-rich data science electives. Hardin et al. (2014) argue that statistics technology is best taught in the context of solving real-world problems as students can see the relevancy and opportunities afforded by technology. Horton, Baumer, and Wickham (2014) discuss strategies for exposing students to precursor data science technology in introductory statistics courses including the use of R for computation (R Core Team, 2014), R Markdown for reproducible analysis (Allaire et al., 2015; Baumer, Çetinkaya-Rundel, Bray, Loi, & Horton, 2014), and R packages for SQL interrogation of databases. Gould and Çetinkaya-Rundel's (2014) innovative "DataFest" has also shed light on important conceptual and technological skill gaps in statistics education. This 2-day, extracurricular competition pits undergraduate student teams against each other in providing insight into real, large, and complex data. Gould and Çetinkaya-Rundel identified key weaknesses in students' ability to generate their own investigations into the data, decide on appropriate units and scales for particular variables, apply multivariate statistical techniques, create visualizations of multivariate data, and analyze spatial and temporal data and general programming in R. Clearly, a modern statistics student needs to be both conceptually and technologically proficient.

Technology may also rapidly change the concepts covered by the introductory curriculum. Perhaps a shift toward accessible forms of statistical inference (Wild, Pfannkuch, Regan, & Horton, 2011) and a resampling-based curriculum (Cobb, 2007) that emphasizes the core logic of inference might achieve technological synergies and a "smarter," more efficient curriculum. Students could be moved further and faster through the current core curriculum and quicker into meaningful, real-world practice. Time could be better spent on building students' capacity to meaningfully engage in real statistical data investigations, performing statistical computing, and working with "big data." This will require a heavy dose of technology.

There are many implications to an increased emphasis on teaching technology within the statistics curriculum. Many challenges must be overcome including addressing poor teacher preparation to use technology (Hassad, 2013; Nolan & Temple Lang, 2010a), an already crowded curriculum facing possible disruption (Nolan & Temple Lang, 2010a), and pedagogical concerns about trade-offs between conceptual understandings (Chance, Ben-Zvi, Garfield, & Medina, 2007). These are legitimate concerns that cannot be brushed aside.

The field of statistics education needs to better understand the relationships between technological skills and statistical knowledge. Are these distinct domains or are these two areas inseparable? A better understanding of the technological skills required to be a successful consumer and producer of statistics, today and into the future, would help shape new and informed curriculum patterns. Statistics educators must look within (Nolan & Temple Lang, 2010a) and outside the field to initiate a coordinated, multidisciplinary, and international effort to teach statistics technology. Patterns of technological skill incorporation and assessment used within statistics education need careful evaluative research in order to ensure students are graduating with the right skills, dispositions, and critical mind-sets to be effective users and producers of statistics technology. This also raises new assessment challenges, because as Gould (2010) reminds us, "We can no longer be complacent and assume students will 'pick-up' the skills they need to negotiate complex data" (p. 309).

15.4 Considering the Gap Between Learning and Doing Statistics

Amelia McNamara
Smith College
Northampton, MA, USA
amelia.a.mcnamara@gmail.com

Most introductory statistics courses now incorporate technology in one way or another. This is a boon to statistics, as true makers of statistical products must use computers. However, there is a distinction between using technology as an instructional aid and fully integrating computation into the curriculum (Biehler, Ben-Zvi, Bakker, & Makar, 2013). Some courses use tools for learning statistics and some tools for doing statistics. Often, this means the difference between students as "creators" of computational statistical work (as discussed by Baglin in this chapter) and as proficient "users" of the technology is not fully engaged in the practice of statistics.

Tools for learning statistics include applets and MicroWorlds—software designed for the purpose of illustrating a statistical concept. Some good examples of applets include those created by Rossman and Chance (Chance & Rossman 2006) and the StatKey applets (Morgan, Lock, Lock, Lock, & Lock, 2014). Students play with applets but cannot modify them beyond the possibilities written in by the author. TinkerPlots and Fathom (Konold & Miller, 2005; Finzer, 2002) provide a landscape-type paradigm (to use a term from Bakker, 2002), but are still designed only for beginners and have limited functionality.

Tools for doing statistics are more powerful and flexible products like R, Python, SAS, and Stata. These tools can be extended (particularly the open-source R and Python) and used to solve a broad range of data problems but are difficult for novices to grasp, particularly when they are also trying to learn statistical concepts.

Tools for learning statistics usually present easy entry for students and tools for doing statistics provide a high ceiling, allowing more flexibility. However, tools that are good for an introductory learner are generally not good for performing real data analysis, and vice versa.

Novices who begin with a tool designed for learning—whether it is an applet or a full software package—must engage in the cognitive task of learning the interface. Creating statistical products, either in higher-level coursework, research, or in the corporate world, requires more complex tools, so learners are forced to learn yet another interface. Researchers have not studied this progression in the context of statistics, so there is little scaffolding to make the transition easier.

Alternatively, novices can immediately start with a tool for doing statistics. This allows them to skip the effort of learning an introductory tool but still incurs the high start-up costs of learning the technology. Whether beginning with a professional tool immediately or transitioning later, the threshold for entry on professional tools can be so high as to make users believe they are not capable of learning it.

It does not have to be this way. As a community, we should consider the learning-to-doing trajectory and develop new ways to support it, by creating new tools and curriculum.

Curricular resources should make explicit reference to the prior tool and couch tasks in terminology from the first system to make transitions easier. Technology should strive to bridge from a supportive environment for learning to an expressive tool for doing. Providing "ramping up" as users reach the end of the abilities of the learning tool and "ramping down" as they start with the tool for doing could make the gap less abrupt.

Since Biehler (1997), researchers have considered what makes an effective tool for teaching statistics (e.g., Ben-Zvi, 2000). In particular, the development of Fathom and TinkerPlots must be commended. Similarly, statistical practitioners are constantly improving the tools used to create true statistical products. For example, Hadley Wickham works tirelessly to produce R packages (ggplot2, dplyr, tidyr) to help practitioners more easily and flexibly produce data analyses (Wickham, 2014).

However, these developments have taken place largely in silos. Where research has crossed over, it has typically been in the context of stripping down professional tools to allow novices to use them. In particular, Project MOSAIC defined a unified R syntax and a limited set of commands to be used in introductory statistics (Horton, Baumer, & Wickham, 2014). This effort, however, considers R as the end goal for students and goes against Konold's philosophy that tools for learning statistics should be developed from the bottom up, thinking about what features novices need to build their understandings (Konold, 2007).

Tools for teaching and doing statistics should aim for the same goals, albeit from different directions. They should strive to provide easy entry for new users while still allowing the flexibility to build extensions onto the system, support a cycle of exploratory and confirmatory analysis, promote interactivity, and make it simple to publish and reproduce analysis. (See McNamara 2015 for further analysis.)

It is time for the research to cross over and for us to think about bridging the gap between students learning statistics and students producing statistical analysis. By developing tools and curriculum to help novices build their understanding from the ground up, we may end up finding better ways for everyone to perform statistical analysis.

15.5 Statistics Education and Empowerment

Jim Ridgway
School of Education
Durham University
DurhamUKjim.ridgway@durham.ac.uk

Statistics emerged as a discipline because people needed tools to improve decision making. Our early history encompasses luminaries such as Nightingale and Fisher, both inventing methods to address practical problems—Nightingale in particular invented tools to change long-ingrained medical practices. The rhetoric of a "useful discipline" persists—the Royal Statistical Society in the UK uses the slogan "data, evidence, decisions." The front page of the International Statistical Institute website is headed *Statistical Science for a Better World.* Effective decision making depends on key statistical ideas such as measurement (including ideas of data provenance and data quality), estimation, probability, utility, expected value, risk, and (of course) different models for decision making. In many countries, the curriculum is dominated by technical mastery of techniques invented in the 1930s to solve a particular, small class of problems—in particular, problems where it is plausible to generalize results from small samples to populations and where rather little data are available to inform some simple decision (e.g., is this crop variety better than that one, in these sorts of conditions?) (Batanero, Burrill, Reading, & Rossman, 2011). Technical mastery of statistical technique is a small part of the decision-making process but is often the focus of both teaching and assessment.

Open data presents more immediate and profound opportunities for curriculum change. There has been a sea change in the ways that evidence is used as a warrant for decision making in the public domain. Open data have profound implications for policy decisions and for the ways that policy is conducted and communicated. Dramatic examples are provided by the 2015 UK national elections, where the claims of politicians about the implications of their proposals (usually in terms of cost) were critiqued by many groups and where real-time fact checking of speeches and debates has become a fact of political life.

Open data have some obvious properties that can be contrasted with data typically used in school classrooms. Data sets are large, so computing power is essential, and almost any comparison between groups will result in a conclusion that differences are "statistically reliable." Hypothesis testing and the determination of *p* values can be set in context as a useful starting point if you are working with modest data sets and small effect sizes. Statistics that are important for decision making, such as bounded estimates of effect sizes, can be taught immediately.

Evidence used in argumentation in the media is often multivariate and assembled from different sources. In our studies (e.g., Nicholson, Ridgway, & McCusker, 2011, 2013; Ridgway & Nicholson, 2010; Ridgway, Nicholson, & McCusker, 2007), young, statistically naïve students do not have problems understanding nonlinearity, interaction, and "third variable" confounding—and, indeed, invent and articulate these ideas for themselves, given interesting data sets to explore (such as data about young people and sex, alcohol, drugs, and rioting). Curriculum practices that build up from basics, and focus on statistical techniques, alienate students for a variety of reasons. It is easy for teachers to say that the subject matter is difficult (it is); more fundamentally, the models implicit in simple linear models of social phenomena are inappropriate, so students see that learning statistics is just a hoop to jump through in order to survive the education process.

When asked to reason with open data related to social issues, students can see that what they are learning can be empowering. The open data movement has been very successful in getting governments and nongovernmental agencies to allow access to data; however, permission to use data does not mean that representative citizens can actually access and use it. Access may require database skills; it is highly likely to require spreadsheet skills and certainly skills in understanding metadata, as well as the skills required to work with tabular data. At school level, students' first experiences should be via rich visualizations of multivariate data on topics they are interested in. At undergraduate level, it is easy to make a case for incorporating data mining, statistical learning, and R and Python into curricula.

A far larger group of people are in urgent need of statistics education, namely, citizens. Data is increasingly being used as evidence in policy making and public debate. How can statistics educators make a difference? My view is that more researchers need to get their hands dirty, by working alongside opinion formers and analysts. Researchers can contribute insights on how to improve comprehension; practitioners can get educators' voices heard and can facilitate public access to new data sources relevant to statistics education. Understanding and influencing the use of evidence in the public domain is the new research horizon. Here is an example.

PARLER is a collaboration between groups at Durham University and the House of Commons Library. The Library is the first port of call for UK politicians who want information about anything—for policy, for an upcoming speech, or to answer a question from a voter. We have created the Constituency Explorer Kit (http://www.constituencyexplorer.org.uk), a collection of linked tools designed to encourage and support exploration of evidence relevant to national elections. In the UK, people are elected from constituencies (geographical areas) to the House of Commons. One component of the Kit is a quiz customized for every constituency that runs on mobile devices; another component is an interactive display which allows users to plot the locus of every constituency on any one of 150 variables; a third component is a visualization that documents changes that arose as the result of the election (again, users can relate changes to a very large number of variables). Display design reflects some important principles from statistics education: Variables can be rescaled, but every time the user chooses a new variable, the scale reverts to 0–100. Display features are explained; variable names are linked to metadata descriptions; data can be downloaded, but metadata must be downloaded, too.

The Library is influential; they work directly with politicians, and their publications, blogs, and press releases are widely used. The Constituency Kit has had a far higher level of exposure that is common in much academic work in statistics education. A major ambition for the collaboration is to create a demand for high-quality data visualizations of key data sources from politicians, journalists, and data suppliers themselves. In this case, an extensive data set relevant to a common theme has been assembled from multiple sources and is universally accessible.

I think that the most important challenge for statistics education is to understand and influence the uses of evidence in decision making. This needs to be an active process—you don't understand a complex system until you try to change it (being bitten by "unknown unknowns" is an important part of learning). Collaborations are essential; working with groups who must be seen to be politically neutral has great advantages. It will not be a research area that can be mapped out much in advance; the systems being researched are changing, and the researchers are part change agents and part careful observers and analysts. Results will feed forward into the actions of organizations that provide data and backward to the statistics curricula that limp along behind.

15.6 Statistics in Journalism: The Present and Possible Futures

Kevin McConway
School of Mathematics and Statistics
The Open University
Milton KeynesUK
kevin.mcconway@open.ac.uk

The fact that statistics plays a role in journalism is hardly news. Even before there was something called data journalism, there were plenty of data in the media. Newspapers were, and are, full of statistical information and of stories based on data—think of the business pages, the sports pages, and even, on the front page, stories based on economic indicators, or health scares based on epidemiological analysis, or body counts from the latest disaster or war.

But what *has* changed is the variety of ways in which data are used and presented. Nowadays, for instance, there is something called data journalism—indeed there are more and more distinctions and variations including data-driven journalism and database journalism. Definitions are fluid, but the basic idea is that methodologies and ways of thinking from other fields that use data are adapted and extended to further the journalist's aim of informing the public.

So journalists might dig through official statistics to find where hospitals are performing badly. Nothing new about the press pointing out that a hospital harmed a patient, but in the past the story would have been about an individual, based perhaps on a court case, rather than on a comprehensive statistical picture. Or it would be based on an official inquiry report where someone else found the data and drew the conclusions.

And, going back a couple of decades, the story would have been in print or on the television, not in an interactive web page that allows the user to look up the exact detail that they are interested in. Some such pages are linked to a more traditional publication such as a newspaper or TV station, but many others are not.

These examples are qualitatively different from "old-style" journalistic outputs and could not exist without the availability of the public data behind them. (Indeed, data journalists have been among the leading advocates within the open data movement.) But even traditional print and broadcast media have changed. There were always a few graphs in the business pages or in articles about the economy, but these days you can scarcely open a newspaper without being confronted with an infographic, usually based on information that is at least partly statistical. Some of these graphics are informative; others seem to be there almost entirely to make the newspaper page look better.

How, then, are all these changes reflected in the training of journalists in statistical matters and the use of data? In the past, the training that a typical journalist had in this area was rudimentary. The journalist Peter Wilby wrote, "Journalists are not very good with figures. The great majority come from an arts or social studies background. […] Most journalism training courses do not have modules on how to handle numbers. Literacy is considered essential for reporters—or at least their subeditors—but not numeracy" (Wilby, 2007).

Things have changed since then, but on the whole not much. For instance, the UK's National Council for the Training of Journalists does now require numeracy skill development in the courses it accredits, but this is recent, and the level of skills required is not high.

It is possible to write an acceptable, or even excellent, news story based on analysis of data without having more than a basic level of these skills, provided one relies on someone else's analysis. Those providing data to journalists often provide such an analysis. In releasing data, government statisticians generally go much further than providing the raw numbers, by presenting them in an analytic report, complete with an executive summary and helpful (to journalists anyway) press releases. A health story based on epidemiological data usually gets into the media because the academic journal, or the university involved, writes a press release, and though journalists would be given access to the paper in the journal, what appears in the news often resembles the press release very closely.

The problem with this journalistic approach to data is that the journalists may not have the skills to ask the right critical questions of the data producers (to say nothing of the fact that they may also not have the time, given editorial demands such as tight deadlines). A 2014 study of science news (Sumner et al., 2014) found that press releases frequently exaggerated the findings of the studies they were supposed to be summarizing, and when they did, news reports tended to cover the story in the same exaggerated manner. This failure to be able to question seems to me to be more prevalent and on the whole more important than simple failings of numeracy.

There have always been journalists in specialist areas who have much stronger numeracy and statistical skills, and these are now joined by data journalism experts who, between them, possess (and need to possess) the range of skills needed in that other loosely defined rising field, data science. Besides statistical skills, they need to have relevant advanced computing skills (including database and Web skills), expertise in design and graphics, as well as the traditional journalistic abilities like recognizing a good story, asking awkward questions, and writing well and persuasively. Training and reference materials specifically for data journalism do already exist—some universities run courses, and there are more widely available sources, such as the online *Data Journalism Handbook* (Gray, Chambers, & Bounegru, 2012) and an introductory MOOC (Massive Open Online Course) run by the European Journalism Centre (2014).

What are the implications for statistics education? A risk is that statistics educators are seen, by journalists, to be irrelevant in this context. It would probably be feasible for the required training of journalists, at all levels, to come from the journalism community. The number of statistically competent journalists may be relatively small at present, but it is not zero and could expand. No, the risk is to the statistics education community, who has important things to learn from the journalists. Statistics educators *are* engaging with journalism educators (see, e.g., McConway, 2015), but this needs to happen more. Statisticians must learn more and more broadly about how and why other professional groups produce and use statistics. We statisticians can't know, and certainly can't control, everything that others get up to with data.

In particular, we must not assume that what we teach about statistics to, say, scientists and social scientists can easily be transferred to teaching journalists. The context is very different. For instance, the European Journalism Centre MOOC (2014) covered the use of Excel spreadsheets to find story ideas. The technical aspects of using Excel were much the same as one might teach in an introductory statistics course for business students. But the instructor pointed out that the news story from a set of data would generally lie in its extreme values, because they are potentially more newsworthy than averages. A statistician would not usually start there. Journalists do still need to know about averages and typical values, but for different reasons.

Journalists are also, typically, under time pressure much more severe than is the case for statisticians and the scientists we often teach. That has implications for what is possible and for how analysis tools must develop—and the implications arguably go beyond journalistic contexts into others where one is data rich but time poor. Journalists need, in most cases, to be good at statistical thinking, not at statistical calculation or at using statistical software. There are parallels with the statistical aspects of teaching of critical appraisal of the literature that medical students (and others) learn, but the details are different and the applicability of the approach is broader.

Finally, a journalist cannot survive without knowing how to tell stories well. The storytelling aspect of statistical analysis and reporting has received attention in the statistics education literature, but we can learn more from journalists on this aspect than they can learn from us.

Statisticians are already involved in training journalists. Let's continue and broaden this engagement. Let's be sure to put ourselves in a position where journalists and statisticians can learn effectively from each other.

References

Allaire, J. J., Joe Cheng, Yihui Xie, Jonathan McPherson, Winston Chang, Jeff Allen, et al. (2015). *rmarkdown: Dynamic Documents for R [R Package]*. Retrieved from http://cran.r-project.org/web/packages/rmarkdown/index.html.

Baglin, J., & Da Costa, C. (2012). An experimental study evaluating error management training for learning to operate a statistical package in an introductory statistics course: Is less guidance more? *International Journal of Innovation in Science and Mathematics Education, 20*(3), 48–67. Retrieved from http://ojs-prod.library.usyd.edu.au/index.php/CAL/article/view/5809/6529.

Baglin, J., & Da Costa, C. (2013a). Applying a theoretical model for explaining the development of technological skills in statistics education. *Technology Innovations in Statistics Education, 7*(2), article 7. Retrieved from http://www.escholarship.org/uc/item/8w97p75s#.

Baglin, J., & Da Costa, C. (2013b). Comparing training approaches for technological skills development in introductory statistics courses. *Technology Innovations in Statistics Education, 7*(1), Article 1. Retrieved from http://www.escholarship.org/uc/item/1hc308sv.

Baglin, J., & Da Costa, C. (2014). How do students learn statistical packages? A qualitative study. In H. MacGillivray, M. A. Martin, & B. Phillips (Eds.), *Topics from Australian Conferences on Teaching Statistics: OZCOTS 2008–2012* (pp. 169–187). New York, NY: Springer Science+Business Media.

Bakker, A. (2002). Route-type and landscape-type software for learning statistical data analysis. In B. Phillips (Ed.), *Proceedings of the 6th International Conference on Teaching of Statistics (CD-ROM)*. International Statistical Institute: Voorburg, The Netherlands.

Batanero, C., Burrill, G., Reading, C., & Rossman, A. (2011). Teaching statistics in school mathematics – challenges for teaching and teacher education. In *Proceedings of the 18th Study of the International Commission on Mathematical Instruction*. Heidelberg: Springer.

Baumer, B., Çetinkaya-Rundel, M., Bray, A., Loi, L., & Horton, N. J. (2014). R Markdown: Integrating a reproducible analysis tool into introductory statistics. *Technology Innovations in Statistics Education, 8*(1). Retrieved from http://escholarship.org/uc/item/90b2f5xh.

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1&2), 127–155.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review, 65*(2), 167–189.

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–690). Springer.

Chance, B., & Rossman, A. (2006). Using simulation to teach and learn statistics. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics (CD-ROM), Salvador, Bahia, Brazil, July 2006*. International Statistical Institute: Voorburg, The Netherlands.

Chance, B. L., Ben-Zvi, D., Garfield, J. B., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1), article 2. Retrieved from http://escholarship.org/uc/item/8sd2t4rr.

Cobb, W. (2007). The introductory statistics course: A ptolemaic curriculum. *Technology Innovations in Statistics Education, 1*(1), article 1. Retrieved from https://escholarship.org/uc/item/6hb3k0nz.

European Journalism Centre (2014). *Doing journalism with data: First steps, skills and tools*. Retrieved from http://datajournalismcourse.net/.

Finzer, W. (2002). Fathom: Dynamic data software (Version 2.1) [Computer software]. Emeryville, CA: Key Curriculum Press.

Gould, R. (2010). Statistics and the modern student. *International Statistical Review, 78*(2), 297–315.

Gould, R., & Çetinkaya-Rundel, M. (2014). Teaching statistical thinking in the data deluge. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen—Using Tools for Learning Mathematics and Statistics* (pp. 377–391). Wiesbaden: Springer Fachmedien Wiesbaden.

Gray, J., Chambers, L., & Bounegru, L. (Eds.). (2012). *The data journalism handbook: How journalists can use data to improve the news*. Sebastopol, CA: O'Reilly Media. http://datajournalismhandbook.org/.

Hardin, J., Hoerl, R., Horton, N. J., & Nolan, D. (2014). Data science in the statistics curricula: Preparing students to "think with data." *arXiv:1410.3127v3.* Retrieved from http://arxiv.org/abs/1410.3127.

Hassad, R. A. (2013). Faculty attitude toward technology-assisted instruction for introductory statistics in the context of educational reform. *Technology Innovations in Statistics Education, 7*(2), article 11. Retrieved from https://escholarship.org/uc/item/9k19k2f7.

Horton, N. J., Baumer, B. S., & Wickham, H. (2014). Teaching precursors to data science in introductory and second courses in statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July 2014. International Statistical Institute: Voorburg, The Netherlands. Retrieved from http://icots.info/9/proceedings/pdfs/ICOTS9_C164_HORTON.pdf.

Konold, C., & Miller, C. D. (2005). TinkerPlots: Dynamic data exploration [Computer software]. Emeryville, CA: Key Curriculum Press.

Konold, C. (2007). Designing a data tool for learners. In M. Lovett & P. Shah (Eds.), *Thinking with data: The 33rd annual Carnegie Symposium on Cognition* (pp. 267–292). Hillside, NJ: Lawrence Erlbaum Associations.

McConway, K. (2015). Statistics in the media: A statistician's view. *Journalism, 17*(1), 49–65.

McNamara, A. (2015). *Bridging the gap between tools for learning and for doing statistics*. Doctoral dissertation. University of California, Los Angeles. Retrieved from https://iase-web.org/documents/dissertations/15.AmeliaAhlersMcNamara.Dissertation.pdf.

Morgan, K. L., Lock, R. H., Lock, P. F., Lock, E. F., & Lock, D. F. (2014). StatKey: Online tools for bootstrap intervals and randomization tests. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics, ICOTS9, July 2014*. International Statistical Institute: Voorburg, The Netherlands.

Nicholson, J., Ridgway, J., & McCusker, S. (2011). Visualise then Conceptualise. *Social Science Teacher, 40*(3), 8–12.

Nicholson, J., Ridgway, J., & McCusker, S. (2013). Getting real statistics into all curriculum subject areas: Can technology make this a reality? *Technology Innovations in Statistics Education, 7*(2), article 1. Retrieved from http://escholarship.org/uc/item/7cz2w089.

Nolan, D., & Temple Lang, D. (2010a). Computing in the statistics curricula. *The American Statistician, 64*(2), 97–107. Retrieved from http://www.tandfonline.com/doi/abs/10.1198/tast.2010.09132.

Nolan, D., & Temple Lang, D. (2010b). Integrating computing and data technologies into the statistics curricula. In C. Reading (Ed.), *Proceedings of the 8th International Conference on Teaching Statistics, July 2010*. Slovenia: Ljubljana. Retrieved from https://iase-web.org/documents/papers/icots8/ICOTS8_9B2_NOLAN.pdf.

Core Team, R. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: The R Project for Statistical Computing. Retrieved from http://www.r-project.org.

Ridgway, J., & Nicholson, J. (2010). Pupils reasoning with information and misinformation. In C. Reading (Ed.), *Proceedings of the 8th International Conference on Teaching Statistics, July 2010*. Slovenia: Ljubljana. Retrieved from http://iase-web.org/documents/papers/icots8/ICOTS8_9A3_RIDGWAY.pdf.

Ridgway, J., Nicholson, J., & McCusker, S. (2007). Reasoning with multivariate evidence. *International Electronic Journal of Mathematics Education (IEJME), 2*(3), 245–269.

Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Venetis, C. A., Davies, A., et al. (2014). The association between exaggeration in health related science news and academic press releases: retrospective observational study. *The British Medical Journal (BMJ), 349*, g7015.

Wickham, H. (2014). *Advanced R*. Chapman & Hall/CRC The R Series.

Wilby, P. (2007. *Damn journalists and statistics*. The Guardian. Retrieved from http://www.theguardian.com/media/2007/nov/05/mondaymediasection.pressandpublishing.

Wild, C., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society, 174*(2), 247–295.

Chapter 16
Design of Statistics Learning Environments

Dani Ben-Zvi, Koeno Gravemeijer, and Janet Ainley

Abstract The goal of this chapter is to draw attention to the need to think about learning environments and their design as a way of considering how sustainable change in the learning and teaching of statistics can be supported. The goal is not to advocate one particular approach to the design of learning environments but rather to raise awareness of the need to consider this lens in statistics education. We first present the rationale for the importance of a focus on learning environments for statistics education. We provide several examples of learning environments that operationalize and integrate various design perspectives and are informed by two theoretical frameworks: social constructivist theory of learning and realistic mathematics education theory. We discuss these examples in a critical way by comparing and evaluating their designs, looking for common threads among them, and develop from them six design considerations for statistics learning environments. Finally, we discuss implications and emerging directions and goals for further implementation and research.

Keywords Learning environment • Statistics education • Learning and teaching of statistics

D. Ben-Zvi (✉)
Faculty of Education, The University of Haifa, Haifa, Israel
e-mail: dbenzvi@univ.haifa.ac.il

K. Gravemeijer
Eindhoven University of Technology, Eindhoven, Netherlands
e-mail: koeno@gravemeijer.nl

J. Ainley
School of Education University of Leicester, Leicester, UK
e-mail: jma30@le.ac.uk

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7_16

16.1 Introduction

Many of the research studies in the learning and teaching of statistics (reviewed by Garfield & Ben-Zvi, 2007; see also chapters in part II of this handbook) suggest innovative approaches that differ from the traditional classroom practices through which most current statistics teachers learned this subject themselves. However, innovation which addresses only one aspect of the pedagogical context, for example, introducing technological tools in teaching when assessment practices remain unchanged, is likely to have only limited impact. This chapter offers starting points of theory and design for deep learning (Sawyer, 2014) of statistics to develop students' statistical reasoning. To do this, we use a *learning environment* perspective to provide a dynamic, holistic, integrated, and multidimensional framework for sustainable educational change in statistics.

The goal of this chapter is to draw attention to the need to think about learning environments and their design as a way of considering how sustainable change in the learning and teaching of statistics can be supported. We provide several examples of learning environments that operationalize and integrate various design perspectives (e.g., Hickey, Kindfield, Horwitz, & Christie, 2003) and are informed by various theoretical frameworks (social constructivist theory of learning and realistic mathematics education theory). We discuss these examples in a critical way by comparing and evaluating their designs, looking for common threads among them. We emphasize that the goal of this chapter is not to advocate one particular approach to the design of learning environments, but rather to raise awareness of the need to consider this lens in statistics education.

A learning environment perspective can guide statistics educators and researchers to view, design, and assess statistics teaching and learning in designed settings, such as classrooms and online courses, as a holistic entity. It can support the intentional transformation of an educational setting based upon conjectures about how the integration of features of the designed setting will support the learning of statistics. Such an entity is a complex and dynamic educational system, composed of multiple factors: key statistical ideas and competencies (content), engaging tasks, real or realistic data sets, technological tools, classroom culture including modes of discourse and argumentation among students and between students and teachers, norms and emotional aspects of engagement, and assessment methods. Integrating all these factors in order to reform the way statistics is learnt and taught is a challenging endeavor. In addition, the broader community (school-level policy makers, local and national authorities, etc.) plays a significant role in the constitution of the learning environment. For example, tensions may arise between required traditional examinations and alternative assessment methods employed in a learning environment or between national curricula and an emergent and dynamic learning trajectory in the learning environment.

New developments in mathematics, statistics, and science education, and more generally in the learning sciences, provide important ideas and practical implications about the design of learning environments (e.g., Bielaczyc, 2006; Collins, 1999; De Corte, Verschaffel, Entwistle, & van Merrienboer, 2003; Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001; Vosniadou & Vamvakoussi, 2006). These developments highlight the value of rethinking what is taught, how it is taught, and how it is assessed (e.g., Bransford, Brown, & Cocking, 2000). The focus in this chapter is on general characteristics of statistics learning environments that need to be examined and integrated in light of these new developments. Our specific objectives are to first present the rationale for the importance of a focus on learning environments for statistics education; we provide a potential framework for considering aspects of statistics learning environments building on social constructivist background theory and the domain-specific theory of realistic mathematics education (RME). We then present three examples of statistics learning environments used in diverse contexts (primary school, lower secondary school, and tertiary education) and develop from them six design considerations for statistics learning environments. Finally we discuss implications and emerging directions and goals for further implementation and research.

16.2 Learning Environments

Reform in statistics education is required and has been sought and evaluated in recent decades (see Chap. 2 of this volume; Cobb, 1992, 1993; Everson, Zieffler, & Garfield, 2008; Garfield, Hogg, Schau, & Whittinghill, 2002; Moore, 1998; Reston & Bersales, 2008). The core idea that underpins this reform is that learning statistics is not about passively acquiring a set of facts and procedures but rather about actively constructing meanings and understandings of big ideas, ways of reasoning, and articulating arguments, dispositions, and perspectives. Unidimensional changes, such as the redesign of particular tasks or aspects of the curriculum, are not sufficient to make extended and sustainable change (e.g., Cuban, 2003; Darling-Hammond, 1997; Kohn, 1999). We recognize that even comprehensive efforts to change several aspects of teaching and learning statistics are not necessarily a key to success (Savelsbergh et al., 2016).

The research literature in statistics education is filled with success stories, which are of importance to the advancement of the field but have not had a major impact on the way statistics is taught in all levels of education. We propose that one of the reasons for this is the lack of a systematic, comprehensive, and integrated approach to design for educational change. We suggest that what is needed is change in a combination of interrelated dimensions (content, pedagogy, space, time, tasks, tools, assessment, classroom culture, etc.) that can bring about significant and sustainable reform in the teaching and learning of statistics by providing a coherent framework in which each dimension operates synergistically with others. Moore (1997) similarly urged a reform of statistics instruction and curriculum based on strong synergies between content, pedagogy, and technology. A learning environment perspective can provide such a framework. One of the major goals of statistics education is to educate critical, independent, and statistically literate learners who are able to study topics of their own interest and become involved in data-based decisions. A learning environment perspective can provide fertile affordances to support learners' growth and development in this direction.

Design dimensions of statistics learning environments that will be considered and discussed in this chapter are based on a number of principles arising from recent research. In particular, we have drawn on research concerning the importance of prior knowledge and preference for depth over breadth (Bransford et al., 2000), the creation of failure-safe learning communities in which students can learn from their successes and mistakes (Bielaczyc & Collins, 1999; Kapur & Bielaczyc, 2012), the nurture and articulation of learners' diverse expertise, encouragement of reflection and feedback (Boud, Keogh, & Walker, 1985), formative assessment (Clark, 2012; Kingston & Nash, 2011), and enculturation into the statistics discipline (Edelson & Reiser, 2006).

Teacher education in statistics is not just about improving teachers' subject knowledge but also about challenging their thinking about the whole process of statistical inquiry as central to statistical thinking and learning (see Chaps. 10 and 14 of this volume; Pfannkuch & Ben-Zvi, 2011). A learning environment perspective can provide a guiding framework for teachers that can support their professional growth in statistics education.

While any setting in which learning takes place can be viewed from a learning environment perspective, we focus now on statistics teaching and learning that occur in the context of designed[1] learning environments[2] (mostly in classrooms and online settings, but sometimes also at home or in the workplace). The use of the metaphor of an *environment* emphasizes that classrooms are interacting social, cultural, physical, psychological, and pedagogical systems rather than a collection of resources, tasks, and activities or a list of separate factors that influence learning. Because of the complex nature of learning environments, successful design requires a working model of how components of the design that help frame forms of student participation and responsibility are collectively constituted and orchestrated (Lehrer, 2009).

To achieve this kind of balance and orchestration, we argue that learning environments must be designed on the basis of learning theories, which can guide the design, help choose between the options, and lead to better achievement of the pedagogical goals. In the next section, we describe two theoretical frameworks that have been commonly used to guide the construction of learning environments.

[1] Learning occurs in a wide continuum of settings from the "designed" to the "ambient" (Kali, Tabak, Ben-Zvi, et al., 2015). On this continuum, this chapter focuses on designed learning environments rather than informal and ambient ways of learning.

[2] Others use the term *learning ecology* instead of learning environment to emphasize that the educational system is always dynamic and emerging rather than a static entity (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Lehrer & Pfaff, 2011).

16.3 Theories that Can Guide the Design of Learning Environments

The roles of theory in design research and in design of learning environments are complex and dynamic (Jonassen & Land, 2012). These vary in levels of generality. From the most general level to the most specific, these include (1) orienting frameworks or background theories; (2) domain-specific instruction theories as frameworks for action; and (3) local instruction theories/humble theories/hypothetical learning trajectories (Prediger, Gravemeijer, & Confrey, 2015). Theories do not provide straightforward recipes for designing effective learning environments. However, they (1) provide a rationale and motivation to use a learning environment approach rather than merely focusing on content, tasks, or what the teachers are doing and (2) can provide considerations, guidelines, and constraints to the practical task of learning environment design (see more on the nature and use of theories in statistics education in Chap. 11 of this volume).

We take social constructivist theory, which is a well-accepted theory in the education community, as our background theory on teaching and learning. This theory requires instructional designers to think through how students construct new knowledge and how the classroom community might interactively constitute instructional tasks. In addition to this general educational theory, one needs a theory that is specific for mathematics education. For reasons we explain below, we choose RME as our domain-specific instruction theory.

16.3.1 Social Constructivist Theory

According to constructivist theory, people learn by *actively constructing knowledge*, rather than by passively receiving knowledge: new knowledge and understandings are based on the *existing knowledge* and *beliefs* one already has and are grounded in our experiences, understandings, and cultural practices (e.g., Cobb, 1994b; Piaget, 1978; Vygotsky, 1978). The thesis that students construct their own knowledge leads to the following questions (Cobb, 1994a): "What do we as educators/instructional designers want the students to construct?" or "What do we want mathematics/statistics to be for them?" and "How do we create a situation in which students construct what we want them to construct?"

When trying to answer the last question, a learning environment perspective suggests that it is not sufficient to design instructional tasks or instructional activities; rather the whole learning environment needs to be considered. Drawing on social constructivism, we may argue that what counts in the learning environment are not just the tasks as such, but the tasks as they are interactively constituted in the classroom. How the tasks are construed depends largely on the classroom social norms, the forms of interaction, and the pedagogical agenda of the teacher. Those in turn are closely related to the learning goals, including a wider goal for how the students understand the nature of statistics.

From our perspective, prioritizing the investigative nature of statistics (see Chap. 4 of this volume), the latter requires specific *classroom social norms*, expecting students to come up with their own questions and solutions and to explain and justify their thinking. Further, it requires appropriate *socio-statistical norms* or beliefs about what it means to do statistics, which concern ways of reasoning and articulation, dispositions, and perspectives. Thus, a social constructivist belief that students construct their own knowledge and our beliefs about what statistics should entail for the students both create the need to think in terms of learning environments and define how we want to shape those environments. (See the Sect. 16.3.2 for an example of a learning environment that uses these social constructivist theory tenets.)

Social constructivism further determines how one thinks about *symbolic representations*. The key here is the distinction that can be made between *inscriptions*—such as marks on paper—and what these inscriptions signify. From a social constructivist point of view, what inscriptions signify is determined by the social practice in which they are used. For example, circles on paper may signify countable objects for students who are participating in a counting activity, while similar circles may signify characters for students participating in a natural language lesson. Thus, from social constructivist stance, establishing social practices in which such inscriptions are produced and used will be a central issue in the design of learning environments.

Several social constructivist theoretical frameworks have been developed to describe learning as *active participation in a community*. Communities of practice (Wenger, 1998), communities of learners (Rogoff, 1994), and knowledge-building communities (Scardamalia & Bereiter, 2014) are three frameworks that have had considerable influence on educational research and practice. Though they may have some nuanced distinctions, they share three fundamental tenets: *activity*, *participation*, and *enculturation*. The active nature of learning is embodied in students' participation in negotiating meanings, developing understanding, evaluating, and orchestrating their own learning in collaborative environments, all with the guidance of an expert teacher (Barron et al., 1998; Ben-Zvi, 2007; Brown & Campione, 1994; Sfard & Cobb, 2014). These forms of participation are, in turn, viewed as processes of enculturation: students assume increasingly central roles in the classroom community and immerse themselves within a culture of learning through which they acquire competence in language, social practices, rituals, and values of the discipline (Barry, 2007). For the classroom community to function effectively, the students and the teacher must negotiate and agree upon standard values and norms that guide and constrain social behavior (Cialdini & Trost, 1998; Hod & Ben-Zvi, 2015). The participation in a classroom community yields not only valued and shared products but contributes to the ongoing development and growth of all members, as they take up and build on each other's knowledge and actions (Rogoff, Turkanis, & Bartlett, 2001).

The implication of a social constructivist stance is that good pedagogical practice consists of designing learning environments that stimulate students to construct knowledge in learning communities. *Statistical inquiry* is one such approach that provides students with many opportunities to participate, think, reason, and reflect on their learning, as well as to discuss and reflect with their peers. A social constructivist perspective on inquiry does not mean that teachers should never tell students anything directly. Rather it means that learning is enhanced when teachers recognize that "teacher telling" does not automatically result in "student knowing" and pay attention to ways in which learners construct knowledge. Monitoring students' changing conceptions as instruction proceeds can provide insights as a starting point for new instruction.

Research does not provide a recipe for designing effective learning environments, but it does support the value of asking certain kind of questions about the design of learning environments and shows their value and success in certain contexts. We argue that the main reason to adopt a learning environment approach is that it appears to be more effective in helping students build a deeper understanding of statistics (e.g., Baeten, Kyndt, Struyven, & Dochy, 2010; Bransford et al., 2000; Cognition and Technology Group at Vanderbilt, 1998; Sawyer, 2014; Sfard & Cobb, 2014).

16.3.2 Realistic Mathematics Education (RME) Theory

According to social constructivism, everybody constructs his or her own knowledge. This puts teachers—and by extension instructional designers—in a difficult position. For how can you ensure that students construct what you want them to construct? Simon (1995) answered this question by proposing a *hypothetical learning trajectory* (HLT): try to anticipate the mental activities of the students when they engage in the instructional tasks under consideration, and relate those mental activities to the learning goal. By developing, enacting, analyzing, and revising HLTs, the teacher can guide the learning process of the students. Teachers can be supported in designing HLTs by being offered prototypical instructional sequences and the local instruction theories underpinning them. These can provide teachers with frameworks of reference for deciding what mental activities to aim for and choosing instructional tasks that may foster these mental activities.

The domain-specific instruction theory for RME offers a theoretical basis for the design of such local instruction theories. The founding father of RME, Freudenthal (1973), argued that students should experience mathematics as a human activity, similar to the activity of mathematicians. While engaging in mathematical activity, they could *reinvent* mathematics (or statistics) with the help of teachers and tasks. In relation to this, he speaks of *guided reinvention*, which, we may argue, is compatible with constructivism, as both Freudenthal and constructivists have in mind students who construct their own mathematics. Over time, those starting points were elaborated in a domain-specific instruction theory, initially formulated by Treffers (1987) and later worked out in the form of instructional design heuristics by Gravemeijer (2004). Those instructional design heuristics encompass *guided reinvention*, *didactical phenomenology*, and *emergent modeling*.

The *guided reinvention* design heuristic asks the designer to develop a potential reinvention route, of which the starting point should be experientially real to the students in that the students would know how to act and reason sensibly in those situations. Freudenthal (1973) pointed out that the designers could look at the history of mathematics/statistics as a source of inspiration (see, for instance, Bakker & Gravemeijer, 2006 who reviewed the historical phenomenology of mean and median). History could tell them which dead ends to avoid and how breakthroughs were achieved. Streefland (1991) added to this the idea of looking at students' informal solution strategies. Students may invent informal solution strategies that show the germs of the applicable mathematics, which could be used as starting points for a reinvention process. Building on Treffers (1987) and van Hiele (1986), we may argue that the learning goals should be framed in terms of networks of mathematical relations. In relation to this, we introduce the notion of *reification*, in which processes obtain an object-like character (Sfard & Linchevski, 1994). The conception of a distribution may, for instance, evolve from the process of organizing measurements within a space of possible outcomes to conceptualizing a distribution as an object with certain characteristics such as shape, center, spread, and skewedness (Bakker & Gravemeijer, 2004).

The *didactical phenomenology* heuristic, also originated by Freudenthal (1983), argues that mathematical thought things such as concepts, rules, and procedures were invented to organize certain phenomena. As examples of a thought thing, we may think of the conception of "the mean," summarizing a set of data in one number, or the conception of a distribution as a more sophisticated way of grasping a data set. The procedure of calculating the mean would offer an example of a procedure type of thought thing. Designers are advised to investigate how the mathematical thought things they are aiming for organize phenomena in applied situations. According to Freudenthal (1983), they can then use that information to create situations in which the need arises to organize phenomena by the very thought things that are to be invented. Related to this, the advice is to explore the variety of situations in which the thought thing is applied in order to create a broad phenomenological base.

The *emergent modeling* design heuristic refers to the roles that models and modeling can play in supporting the reinvention process. Of key importance here is the notion that symbolic representations do not come with an inherent meaning. In relation to this, Bereiter (1985) framed the learning paradox: to come to understand a new piece of mathematics, you have to understand the symbolic representations that derive their meaning from the very piece of mathematics you want to come to understand. The emergent modeling heuristic aims at circumventing this learning paradox by fostering a learning process in which symbolizations and meaning coevolve. Initially, the models come to the fore as context-specific models, referring to situations that are experientially real for the students. Initial models have to allow for informal solution strategies at the level of the contextual problem. Then, while the students gather more experience with similar problems, the teacher will support them in shifting their attention toward the mathematical relations and strategies. This will help them to further develop those mathematical relations, which enables them to see the model in a different light; the model starts to derive its meaning from the emerging network of mathematical relations and starts to become a base for more formal mathematical reasoning. Thus, a *model of* informal mathematical activity develops into a *model for* more formal mathematical reasoning, together with the development of a network of mathematical relations, which the students may experience as an expansion of their mathematical reality.[3]

Taken together, social constructivism and RME can provide a conceptual foundation to guide the design of learning environments. Although not the only relevant theories, they provide one example of how theory and practice are strongly linked and can enrich each other. We turn now to describe three examples of statistics learning environments which were designed based on one of these theories, each targeting a different age level: primary school, lower secondary school, and tertiary level.

16.4 Examples of Statistics Learning Environments

16.4.1 Example I: The Connections Learning Environment (Primary School)

The connections learning environment is built upon the principles of social constructivist theory (Sect. 16.3.1 above) and is designed for young learners (ages 10–12). It is a design and research project which started in 2005 (Ben-Zvi, Gil, & Apel, 2007) to develop students' statistical reasoning in an inquiry and technology-enhanced learning environment in primary schools in Israel.

The project extends for 5 weeks (6 h per week) each year in grades 4–6 during which students actively experience some of the processes involved in data-based inquiry, mirroring the practice of statistics experts. Students conduct data and statistical modeling investigations through peer collaboration and classroom discussions using TinkerPlots (version 2; Konold & Miller, 2011), a computer tool for dynamic data and modeling explorations. By playing a role in helping students learn new ways of representing data and develop statistical reasoning, TinkerPlots gradually becomes a thinking tool for these students; it scaffolds their ongoing negotiations with data, statistical ideas, inferences and their meanings (Ben-Zvi & Ben-Arush, 2014).

The tasks undertaken by connections students are a series of open-ended real data investigations that provide students with rich and motivating experiences in posing statistical questions; collecting, representing, analyzing, and modeling data; and formulating inferences in authentic contexts, which result in meaningful use of statistical concepts (Ben-Zvi, Aridor, Makar, & Bakker, 2012; Makar & Ben-Zvi, 2011). The data are obtained from a questionnaire designed by the research team, teachers and students, and administered by students in their school. The connections classroom is conceptualized and organized as a learning community (Bielaczyc & Collins, 1999) that supports collaboration, argumentation, sharing, and reflection.

[3] In practice, "the model" in the emergent modeling heuristic is actually shaped as a series of consecutive sub-models that can be described as a cascade of inscriptions or a chain of signification.

This is done physically in the class and virtually in a website that includes all educational materials and supports, students' reflective diaries, and peer and teacher feedback.

Alternative methods of assessment (Garfield & Ben-Zvi, 2008, pp. 65-89) are used as an integral part of the connections learning environment. These assessment activities, including student projects, oral presentations, and peer and teacher feedback, are viewed as an important component of the learning processes rather than only as a means for "testing" of students' outcomes. For teachers, they provide opportunities to gain insights into students' developing constructions of meaning and so are a crucial part of the planning and design process. Students are usually highly motivated to present and discuss their work in short presentations during the project and at the statistical happening, a final festive event with their parents.

In the connections learning environment, rather than first teaching students directly about statistical concepts and then asking them to apply them in investigations, the investigations themselves are designed to raise the need to attend to these concepts, hence deepening students' understanding of both their relevance and application. Additional strategies are used in the design of the educational materials such as growing samples (Bakker, 2004; Ben-Zvi, 2006; Konold & Pollatsek, 2002), which is a pedagogical heuristic in which students are gradually introduced to increasing sample sizes that are taken from the same population. For each sample, they are asked to make an informal inference (Chap. 8 of this volume) and then predict what would remain the same and what would change in the following larger sample. Thus, students are required to reason with stable features of distributions or variable processes and compare their hypotheses regarding larger samples with their observations in the data. They are also encouraged to think about how certain they are about their inferences. Beginning with small samples, students are expected to experience the limitations of what they can infer about this current sample. This is a useful pedagogical tool to sensitize and slowly introduce students to the decreasing variability of apparent signals in samples of increasing sizes.

Ben-Zvi (2006) found that the growing sample heuristic combined with "what if" questions not only helped connections students make sense of the data at hand but also supported their informal inferential reasoning by observing aggregate features of distributions, identifying signals out of noise, accounting for the constraints of their inferences, and providing persuasive data-based arguments. The growing awareness of students to uncertainty and variation in data enabled students to gain a sense of the middle ground of "knowing something" about the population with some level of uncertainty and helped them develop a language to talk about the gray areas of this middle ground (Makar, Bakker, & Ben-Zvi, 2011).

The growing sample heuristic does not stand alone but is part of the broader connections learning environment. For example, the students have a deep grasp of the sample data since they have collected them, and the technological tool allows them to present the growing samples easily and creatively in a supportive classroom culture.

The connections project was based initially on an exploratory data analysis (EDA) pedagogic approach (Shaughnessy, Garfield, & Greer, 1996). Students drew informal inferences from real samples following the statistical inquiry cycle. To foster students' appreciation of the power of their inferences, a model-based perspective has recently been added in which students build a model (a probability distribution) for an explored (hypothetical) population and produce data of generated random samples from their model using TinkerPlots. By analyzing generated random samples and comparing them with the suggested model, students learn about the relationships between samples and populations as well as about sampling variability and representativeness (Manor & Ben-Zvi, 2017).

Connections students gain considerable fluency in techniques common in exploratory data analysis: the use of statistical concepts, statistical habits of mind, inquiry-based reasoning skills, norms and habits of inquiry, and TinkerPlots as a tool to extend their reasoning about data (e.g., Ben-Zvi et al., 2012; Gil & Ben-Zvi, 2011). In a longitudinal mixed method study (Gil & Ben-Zvi, 2014), evidence of long-term impact of teaching and learning was sought among ninth graders, 3 years after their participation in the 3-year connections intervention. Students from two groups, who had/had not taken part in the program, were compared throughout three extended open-ended data inquiry tasks and took a statistical reasoning test. Connections students had significant gains in terms of their conceptual understanding of aggregate view of a distribution and informal statistical inference. They used statistical concepts in a more meaningful, integrated, and accurate manner in their explanations, were more fluent considering the uncertainty involved in generalizations from random samples, and supported their inferences with data-based evidence.

In sum, connections students learn by actively constructing knowledge of key statistical ideas and competencies; enjoy open, extended, and engaging tasks; investigate real data sets with sophisticated technology; and are assessed with alternative methods. The combination of these activities and entities, coupled with supportive and caring classroom learning culture, creates a learning environment that nurtures students' deep statistical learning (e.g., Aridor & Ben-Zvi, 2017; Manor & Ben-Zvi, 2017).

16.4.2 Example II: The Nashville Data Analysis Project (Lower Secondary School)

In the late 1990s, two extended data analysis teaching experiments were carried out in a 7th and an 8th grade classroom. These design experiments (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) were part of a 10-year collaboration of Cobb, Gravemeijer, Yackel, and others, in which RME theory was elaborated while adopting the collectivist perspective on teaching and learning that is implied by a social constructivist view (Cobb, Gravemeijer, & Yackel, 2011). Similar to Freudenthal's (1973) adage of *mathematics as an activity*, the starting point for the design was that students would have to experience learning about data analysis by doing data analysis. This of course required a matching classroom culture in which the teacher and students could function as a community of practice/learning community. The structure of the lessons was tailored to this idea, which started with a whole-class discussion in which a problem or issue was explored, followed by small groups solving the problem with the help of a computer tool, and concluded with presentations and discussions.

This approach was worked out in the following manner (see also Gravemeijer & Cobb, 2013). The tasks were designed on the basis that the students would be doing data analysis "for a reason": to solve a problem or to answer question, preferably concerning a topic that the students considered relevant. To foster an effective reinvention process, a shift was made during the first teaching experiment from solving problems to considering and improving ways of data analysis and visualization, denoted as *cultivating statistical interest*. To achieve this, the students were given the role of data analysts working in the service of people who had to make decisions. The tasks usually involved comparing two data sets. Faithful to the idea of data analyses for a reason, the students were involved in the process of data creation. However, as assembling data was not feasible in most cases, this took the shape of *talking through the process of data creation*. In this manner, the researchers also tried to ensure that the starting points would become experientially real for the students.

Following the emergent modeling design heuristic, the researchers tried to provide for a process in which the ways of symbolizing/visualizing and the development of meaning coevolved. The backbone of the instructional sequence was formed by a series of visual representations that functioned as sub-models in an emergent modeling process (Fig. 16.1). We will briefly describe how this series of submodels evolves.

The starting point is the supposition that 7th grade students will be familiar with representing individual measurement values as lengths. When comparing data sets, the focus of the students will be on the end points of the individual value bars and the corresponding positions on the *x*-axis (Fig. 16.1a). So the bars can be left out, while the end points descend to the horizontal axis, resulting in a dot plot (Fig. 16.1b). Analyzing and comparing distributions represented by dot plots, students may start to reason about the shape of the distribution (Fig. 16.1b). In doing so, the vertical axis will come to signify the density of data points around a given *x*-value. While structuring data sets in various ways, structuring data in four equal groups may come to the fore as one of the powerful ways of structuring data (Fig. 16.1c, d). Here students may start using the partitioning in halves and quarters as means for comparing data sets while also starting to get a handle on distributions by realizing that the data density is the highest where the distance between the vertical bars is the smallest. Then the students may start to use four equal groups or boxplots as means to reason about distributions (Fig. 16.1e, f). Ideally the boxplots will come to signify the shape of the distribution for the students, thanks to the history of its emergence. In the process, distributions are expected to acquire an object-like quality for the students, objects with characteristics such as shape, spread, and skewness—which can be further defined with median, quartiles, and extreme values.

Building on this model, bivariate data sets may be sliced into a series of univariate distributions that can be represented as a series of (vertical) boxplots. Thinking of hill-type shapes corresponding with these boxplots, a ridge can be imagined running across the data set. This ridge may be interpreted in terms of a conjectured relationship of covariation between the two variables involved.

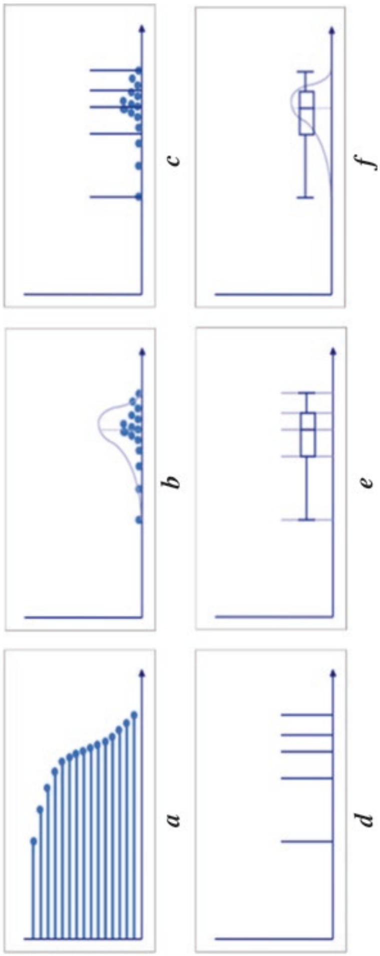


Fig. 16.1 **(a–f)** Sub-models in an emergent modeling process

The visualizations are embedded in computer tools; and the computer tools, with the built-in tool options, were so designed that they would support the aforementioned reflexive process. The first tool shows data values as bars with a dot at the end. The students can structure data in various ways while comparing two or three data sets. The second tool shows data points in the form of dot plots, which can be structured in various ways, in particular in either two or four equal groups. The third tool shows bivariate data sets in a Cartesian graph and allows for slicing the data set vertically and structuring those slices in two or four equal groups.

There were several indicators that the students were in fact reinventing elementary statistics. At the end of the 7th grade, students used the tool options in original ways and invented idiosyncratic concepts such as "consistency" (small spread), "majority" (highest density), and a hill metaphor, which not just signified the visual shape for the students but also the way the data were distributed (Cobb, McClain, & Gravemeijer, 2003; Gravemeijer, 2002a, 2002b). They realized that a higher point on the hill corresponded with a higher density of data points.

At the end of the 7th grade teaching experiment, most students could readily interpret graphs of data sets in terms of characteristics of distributions while focusing on informative ways of organizing data. A limitation, however, was that they did not see the median as a characteristic of the data set, probably due to the fact that the median and the quartiles were often used to partition the data sets in order to compare them multiplicatively. However, the students did develop the notion of "hill" and "majority," which later on (in the 8th grade) could be further developed into the interpretation of the median as indicator of the location of a hill.[4]

16.4.3 Example III: Statistical Reasoning Learning Environment (Tertiary Level)

Garfield and Ben-Zvi (2008, pp. 45-64; 2009) designed a learning environment model for an interactive, introductory secondary- or tertiary-level statistics course that is intended to develop students' statistical reasoning. This model is called a "Statistical Reasoning Learning Environment" (SRLE) and is built on the social constructivist theory of learning (Sect. 16.3.1 above). The model is also recommended for use in teacher education (Pfannkuch & Ben-Zvi, 2011).

The SRLE may be better understood through comparison with a "traditional" university class. In a "traditional" class, the students come to class with no anticipation of what they will learn, ready to copy down what the teacher has to say. The teacher presents a lecture that includes examples, some data analysis, and perhaps some demonstrations. The students listen, take notes, and perhaps ask questions. They leave the class with a homework assignment that uses information from the class they just attended. They go home and try to solve the problems by looking

[4] Note, however, that the students have to be made aware that not all distributions are unimodal.

back at their notes or looking up worked examples in the textbook, often getting frustrated if they do not find an exact match.

In an SRLE class, the students know that they have to prepare for class by reading a few pages in the textbook using study questions to guide their reading and note taking or by responding to a task, such as a data analysis task or an interview with a child. Students therefore come to class with a preliminary exposure to the topic, and sometimes with questions about it. Class begins with a short summary of, and reflection on, what was learned in the previous class, and students are asked if they have questions about the previous class or the assigned task. Students ask some questions that are answered by other students and/or the teacher. The teacher rarely answers a question directly but often asks students, "What do you think?," and if another student gives an answer, the teacher asks, "Do you agree with this answer? Why?"

Now the class is ready to begin the first task. A question is given to the students such as "Do you think that female students spend more time on cell phones than male students?" Students form small groups to discuss these questions and sketch possible distributions and then share and compare their conjectures and reasoning with the class. The students move to computers and access a data set containing information that has previously been gathered about the students in the class using an online student survey. Working in pairs, students generate graphs and statistics to answer the questions on cell phone use. Students may discuss appropriate measures of center and spread for the data, revisiting those ideas from previous lessons. They may notice outliers in the data, and discussion may focus on how to find out if these are legitimate values or errors and on what happens to the graphs and statistics if those extreme values are removed?

The teacher's role in the SRLE class is to present the problem, guide the discussion, anticipate misconceptions or difficulties in reasoning, and make sure students are engaged on task and not experiencing difficulties. The teacher has to know when to end discussions, how to learn from mistakes, and how to provide a good summary for the task using the work students have done, so students can appreciate what they learned from the task. At the end of class, after the wrap-up discussion and summary, students may be asked to complete a brief assessment task, providing the teacher with feedback on their learning for that class.

The contrast between the SRLE and traditional instructional approaches is large, and it is apparent that even an eager and enthusiastic teacher who wants to move from a more traditional approach to a more SRLE approach is faced with many challenges. These challenges include students, colleagues, and institution, as well as challenges to the teacher's own expectations. These challenges are examined and addressed in Garfield and Ben-Zvi (2008, pp. 57-63).

The SRLE model integrates many previous research results and is based on current learning theories. It is hard to envision a way to empirically test it in its entirety since it is too complex and could translate differently in different courses and educational levels. Indeed, there is little empirical evidence as to what extent the entire SRLE improves students' statistical reasoning and thinking (Baglin, 2013; Loveland, 2014). Conway (2015) studied the impact of conformity to SRLE principles on students' statistical reasoning in advanced placement statistics courses[5] in the USA. While the comparison between classrooms showing low and high conformity to SRLE principles revealed no statistically significant differences in students' statistical reasoning ability, results from this study suggest that beliefs and practices aligned with SRLE principles show potential to increase students' statistical reasoning at rates above national averages.

Several aspects of the SRLE were studied to assess learning outcomes. For example, both cognitive and affective/motivational factors were found associated with using real-life data to teach statistics in a first-year university statistics course (Neumann, Hood, & Neumann, 2013). Slootmaeckers, Kerremans, and Adriaensen (2014) used similar principles in the integration of quantitative material into non-methodological courses for political science students. Their results indicate that such an approach can not only foster interest in statistics but also retention of the acquired statistical skills.

16.5 Design Dimensions for Statistics Learning Environment

In this section, we identify design dimensions that arise from theoretical and empirical sources we have discussed and the three learning environment examples described in the previous section. These dimensions are not meant to serve as a prescription for what teachers and designers should do but rather to provide a wide spectrum of factors, or starting points, that need to be considered, aligned, and balanced in designing statistics learning environments. The goal of designing effective and positive statistics learning environments is to support students to develop a deep and meaningful understanding of statistics and the ability to think and reason statistically. In considering the design of such learning environments, we discuss and expand on six dimensions of pedagogical design proposed by Cobb and McClain (2004), highlighting what we see as the important connections between them (Fig. 16.2).

16.5.1 Focus on Developing Central Statistical Ideas Rather than on Tools and Procedures

There are several key statistical ideas that school and university students are expected to understand at a deep conceptual level (Burrill & Biehler, 2011; Garfield & Ben-Zvi, 2008). These ideas serve as overarching goals that direct teaching and

[5] Advanced placement is a US academic program with more than 30 courses in a wide variety of subject areas that provides secondary school students with the opportunity to study and learn at the college level.

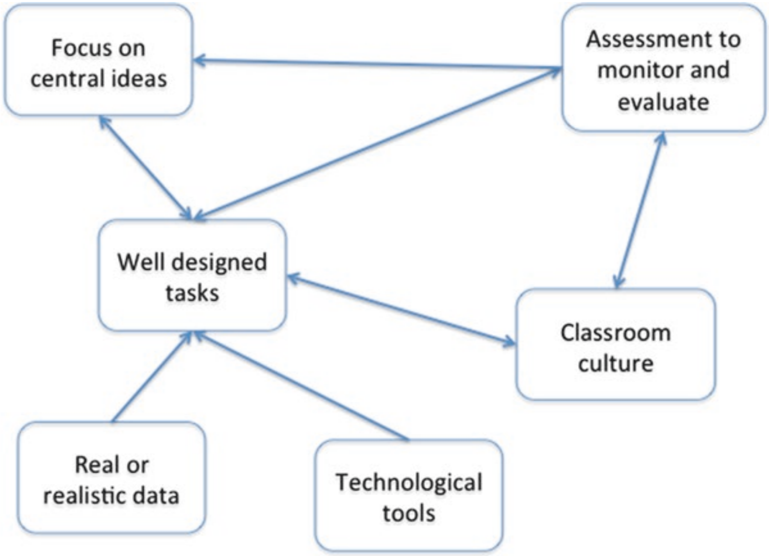


Fig. 16.2 A web of interrelated dimensions of a learning environment

motivate and guide students' learning. They include data, distribution, center, variability, comparing groups, sampling, modeling, inference, and covariation.

Garfield and Ben-Zvi (2008; see Example III above) advocate a focus on key statistical ideas and the interrelations among them and suggest ways to present these ideas throughout a course, revisiting them in different contexts, illustrating their multiple representations and interrelationships, and helping students recognize how they form the supporting structure of statistical knowledge.

Following the RME approach, one would aim for reinventing statistical ideas and allowing procedures and definitions to emerge in the process of coming to terms with a key idea. As exemplified earlier with the example of the process of reinventing the conception of distribution as a mathematical object, measures of central tendency may be developed as means to get a handle on distributions.

16.5.2 Use Well-Designed Tasks to Support the Development of Statistical Reasoning

An important part of a statistics learning environment is the use of carefully designed tasks, informed by research findings, that promote student learning through collaboration, interaction, discussion, and addressing interesting problems (e.g., Roseth, Garfield, & Ben-Zvi, 2008). It may be argued that such tasks should be part of a well-considered instructional sequence, informed by the aim of developing central statistical ideas, which is underpinned by a *local instruction theory*. A local instruction theory typically consists of a theory about a potential learning process and theories about the means of supporting that learning process (Gravemeijer & Cobb, 2013). The former offers teachers background information, on the basis of which they may decide, on a daily basis, what learning goals to aim for, while the latter offer them information on how potential tasks, tools, ways of interacting, and the classroom culture may support the intended learning process. This information will help teachers in choosing tasks and tools, anticipating the mental activities of the students, orchestrating classroom interaction, and evaluating the implied hypothetical learning trajectories.

Anticipating the notion of density, for instance, a step in the learning process, will concern the shift from measures represented as proportionally seized horizontal bars to measures represented by dots—where positions of the dots correspond with the end points of the bars (see Fig. 16.1a, b). The key here is to orient the students toward the end points of the bars and their position in respect to the horizontal axis. This asks for tasks in which the positions of the end points of the bars form a central issue. The battery life span task (Fig. 16.3) nicely fulfills this requirement (although the teacher may choose another task that can fulfill this function).

Designing high-quality tasks is demanding, not least because of some inherent tensions. One of these, which Ainley, Pratt, and Hansen (2006) call the "planning paradox," is between engaging students' interest by allowing freedom for them to develop their own ideas and ensuring that specific mathematical or statistical ideas are addressed. This is a tension between the design of appropriate tasks and the constraints of the institutional learning context.

Addressing this challenge, Ainley and Pratt (2014a) propose two linked principles for task design which are particularly appropriate in statistics education and potentially have wide application within the design of learning environments. The first is that tasks should have a clear *purpose* for the students within the context of the classroom. This might involve making a real or virtual object, such as a paper spinner or a model to generate data, or solving an intriguing problem. The purpose in this sense is not necessarily related to a real-world application: the purpose may arise within a fictional context, such as students advising on the movement of a new character in the "Angry Birds" computer game (Ainley & Pratt, 2014b). What is important is that the challenge of the task is engaging for students.

The second principle concerns the *utility* of statistical ideas, that is, the ways in which these ideas are useful. Engaging tasks should offer students opportunities to use statistical ideas in ways that enable them to see how and why they are powerful. For example, in modeling the movement of an "Angry Bird" which only moves horizontally (an "Angry Emu"), students can appreciate the need to express both signal and noise to describe the distance the Emu will travel relative to how far the sling is pulled back.

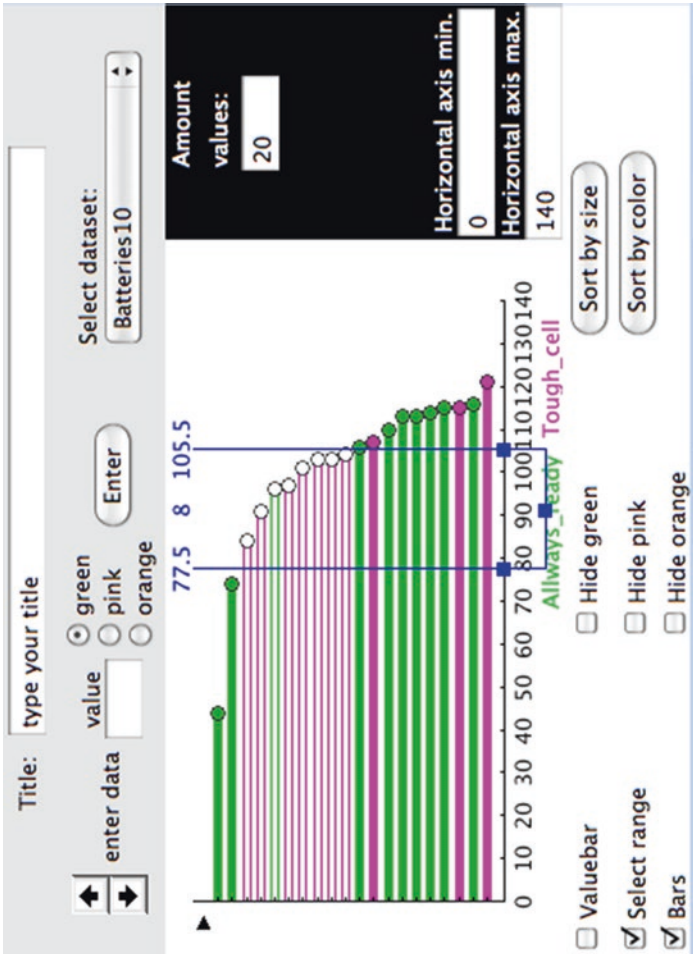


Fig. 16.3 Comparing data on the life span of batteries with a computer tool

16.5.3 Use Real, or Realistic, and Motivating Data Sets

The design of pedagogic tasks in statistics must take account of the data that will be centrally involved. Data are at the heart of statistical work, and data should be the focus for statistical learning as well (Franklin & Garfield, 2006). Throughout their experience of learning statistics, students need to consider methods of data collection and production and how these methods affect the quality of the data and the types of analyses that are appropriate. One approach can be to look for interesting data sets to motivate students to engage in activities, especially ones that ask them to make conjectures about a data set before analyzing it (Ben-Zvi & Aridor, 2016). Another approach would be to start with a question and then discuss what data would be needed to answer it. However, the provision of real or "realistic" data is not always sufficient to engage students in tasks that develop statistical reasoning unless the task poses meaningful challenges and provides opportunities to use statistical ideas in realistic ways.

Consider two kinds of activities using real data which are relatively familiar within statistics education research. The first is exploratory data analysis based on a source of real data, such as CensusAtSchool (Connor, 2002). Although data about students like themselves may have intrinsic interest, posing meaningful questions about the data can be challenging for school students (e.g., Burgess, 2007). Open-ended exploration of relationships in the data without a clear goal may not lead them to use statistical ideas in realistic ways. The second is a sampling task, such as repeatedly drawing small samples to estimate the proportion of sweets of a particular color within a bowl. Here, the statistical idea of sampling is being used in a realistic way, to answer a specific question, but the task itself is not a meaningful challenge (Ainley, Gould, & Pratt, 2015). If you really wanted to know the numbers of sweets of different colors, it would be quicker and more reliable to empty the bowl and count them. What these tasks have in common is that, although based on real data, they do not emphasize opportunities for students to appreciate the utility of statistical ideas. As a result, they may appear contrived and fail to engage and motivate students.

There is a further tension concerning the role and nature of data in statistics tasks. Students, particularly younger students, need to experience collecting, recording, and cleaning their own data in order to develop their understandings of different forms of representation (e.g., Neumann et al., 2013). But data collection is time-consuming, often leaving relatively little time for analysis and discussion, and the features of the resulting data sets cannot be predicted. Providing real-world data sets (such as CensusAtSchool data, e.g., http://new.censusatschool.org.nz/), or devising data sets which are not authentic but embody the features that the teacher wants students to experience, will save time, but students may find such data sets harder to understand and engage with (Arnold, 2014).

In their teaching experiment on data analysis, Cobb et al. (2003) used an approach, which offers a resolution to this tension. They asked seventh grade students what data would be needed for a consumer report on batteries, providing an overall purpose for the task. The students came up with the variable "life span" and figured out how data on life span would have to be gathered, in an activity which the authors describe as "talking through the process of data creation." Subsequently, the students were offered life span data on two brands of batteries, which were not authentic but tailored to focus attention on statistical ideas involved in comparing two data sets, as part of instructional sequence. The data sets were constructed in such a way that one data set had a number of long-lasting batteries, while the other data set had a smaller spread. This allowed for a discussion of a small spread, for which the students invented the term "consistency" versus some high values. Students eventually linked this to the issue of what you would want to use the batteries for.

16.5.4 Integrate the Use of Technological Tools that Allow Students to Explore and Analyze Data

The design of tasks (Watson & Ohtani, 2015) and the ways in which students may access and explore data are significantly influenced by the range of technological tools available to support the development of students' understanding and reasoning, such as computers, graphing calculators, Internet, statistical software, and web applets (e.g., Biehler, 2003). Students no longer have to spend time performing tedious calculations, or drawing graphs, and can focus instead on the more important task of learning how to choose appropriate analytic methods and how to interpret results. Technological tools are used not only to generate statistics, graph data, or analyze data but also to help students visualize concepts and develop an understanding of abstract ideas through simulations. For examples of innovative tools and ways to use these tools to help develop students' reasoning, see Ben-Zvi (2000); Chance, Ben-Zvi, Garfield, and Medina (2007); and Biehler, Ben-Zvi, Bakker, and Makar (2013).

A special category of technological tools is that of tools that are tailor made to instructional sequences, aiming to support "guided reinvention." As an example, we may refer to the data analysis experiment of Gravemeijer and Cobb (2013) described above that aimed to develop students' understanding of distribution as an object. Here an emergent modeling approach was applied in which the various sub-models instantiated the overarching idea of visualizing data sets. These visualizations were embedded in computer tools which enabled the students to structure the data in various ways. When comparing two data sets on the life span of batteries, for instance, the students used the tool options to compare the values of the AlwaysReady batteries with those of the Tough Cell batteries (Fig. 16.3).

Referring to the computer tool representation, they argued that they would prefer the "consistency" of the Tough Cell batteries over the many high values of the AlwaysReady batteries, when they needed a battery to really rely on; it would give you at least 80 h.

16.5.5 Establish a Classroom Culture that Fosters Statistical Arguments

The design of tasks and technological and assessment tools has to take into account the expected forms of classroom *discourse*. In statistics learning environments, the use of activities and technology allows for a form of classroom discourse in which students learn to question each other and respond to such questions, as well as explaining their answers and arguments. Cobb and McClain (2004) describe the effective classroom discourse in which statistical arguments explain why the organization of data gives rise to insights about the phenomenon under investigation and students engage in sustained exchanges that focus on significant statistical ideas.

It can be challenging to create a statistics learning environment with classroom discourse that enables students to engage in discussions in which significant statistical issues emerge and where arguments are presented and their meaning is openly negotiated. Creating a classroom climate where students feel safe expressing their views, even if they are tentative, is another challenging task and is related to classroom *culture*, in which the teacher and students have to develop the corresponding classroom social norms and socio-mathematical (or socio-statistical) *norms* (Yackel & Cobb, 1996). These norms encompass the obligation for the students to explain and justify their solutions, to try to understand the explanations and reasoning of the other students, to ask for clarification when needed, and eventually to challenge the ways of thinking with which they do not agree. The teacher is not expected to give explanations but to pose tasks and ask questions that may foster students' thinking. Socio-statistical norms would be tailored to what it means to do statistics, for example, what a statistical problem is, what a statistical argument is, and so forth.

As described in the three learning environment examples above, the shift in the classroom culture is related to a potential shift in the role of the students, from problem-solvers to statisticians who analyze and represent data to make them easily accessible for decision makers. When adopting the role of a data analyst, or data detective (Pfannkuch & Rubick, 2002), students can start reflecting on the adequacy and clarity of condensed descriptions and representations of data, which may foster the reinvention of more sophisticated representations and concepts.

16.5.6 Use Assessment to Monitor the Development of Students' Statistical Learning and to Evaluate Instructional Plans

Assessment should be aligned to well-designed tasks that focus on central statistical ideas in a discourse-rich classroom. Much of the value of changes in the other design dimensions will be lost if assessment practices are not aligned in this way, since the attention of students and teachers will be shaped by the requirements of assessment. In recent years, many alternative forms of assessment have been used in statistics classes. In addition to quizzes, homework, and exams, many teachers use statistical projects as a form of assessment (MacGillivray & Pereira-Mendoza, 2011). Other forms of alternative assessment are also used to assess students' statistical literacy (e.g., critique a graph in a newspaper) and reasoning (e.g., write a meaningful short essay) or to provide feedback to the teacher (e.g., minute papers) (Bidgood, Hunt, & Jolliffe, 2010; Franklin & Garfield, 2006; Gal & Garfield, 1997).

Assessments need to be aligned with learning goals, focusing on understanding key ideas and not just on skills, procedures, and computed answers. This can be done with formative assessments used during a course (e.g., quizzes, small projects, or observing and listening to students in class) as well as with summative evaluations (course grades). Useful and timely feedback is essential for assessments to lead to learning. Types of assessment may be more or less practical in different types of courses. However, it is possible, even in large classes, to implement good assessment practices (Garfield & Ben-Zvi, 2008, pp. 65–89).

16.6 Discussion: Contemporary Issues and Emerging Directions

The goal of this chapter has been to draw attention to the need to think about learning environments and their design in statistics education as a way of considering how sustainable change in the learning and teaching of statistics can be supported. It is not to advocate one particular approach to the design of learning environments, but rather to raise awareness to the need to consider this lens in statistics education research and practice. We have provided several examples of statistics learning environments that were informed by the social constructivist and the realistic mathematics education theories. Drawing on these examples and theories, we have discussed six dimensions of statistics learning environments.

Designing for educational change to support the development of students' statistical reasoning is a challenging task. Using a lever to make a one-dimensional change (e.g., formulate new tasks, the use of a new pedagogical strategy) may make a difference that is not necessarily a sustainable change in students' understanding of statistical ideas. This chapter has argued for a holistic and integrated approach that advocates a learning environment where students are engaged in making and testing conjectures using data, discussing and explaining statistical reasoning, focusing on the important big ideas of statistics, using innovative tools in creative ways to assist their learning, and being assessed in appropriate ways.

We have discussed how the design of a statistics learning environment might take into consideration the following interrelated dimensions: a focus on central statistical ideas, the use of real or realistic data sets, well-designed tasks, integration of the use of appropriate technological tools, promoting classroom culture that nurtures discourse and socio-statistical norms, and the use of appropriate assessment methods (Cobb & McClain, 2004).

A key factor in this discussion is that these dimensions, which are interrelated (see Fig. 16.2), must be aligned and balanced. Issues of alignment are important for accelerating statistics learning both within and outside of schools. The meaning of these design principles being part of integrative whole is that using one of them separately is not enough to make deep and sustainable change in students' learning. The learning environment approach helps to interlink them. For example, the design of motivating tasks is linked to real data collection; these data can be used to build students' statistical understanding taking advantages of the innovative affordances of technological tools; productive classroom discourse is supported by the design of open-ended tasks that support argumentation and by appropriate responses by the teacher (Makar, Bakker, & Ben-Zvi, 2015); assessment methods need to align with the design of tasks; a provision of a new tool must consider the potential interactions with content and pedagogy (Moore, 1997).

Thus we argue that pedagogical and research efforts for change must consider the interactions among these dimensions. There are however other important dimensions of learning environments that were not included in this chapter. One example is the emotional aspects of engagement and identity to motivate all students to participate and reflect on their experiences (Heyd-Metzuyanim, 2013).

Learning environments should become part of the statistics education community discussion. Rather than the limited current focus on a specific tool or a set of innovative tasks, we hope to see more studies that report on integrated learning environments in statistics. The challenge is manifold. Planning a learning environment study is more complicated than a single-factor experiment, there are possibly greater tensions with local and national institutional constraints, and the design of assessment has to take into account multiple dimensions and use mixed methods.

If taken seriously, there are contemporary issues and future directions in this area of statistics learning environments. First, further research is crucially needed to provide more well-researched holistic examples in different contexts and age levels. Systematic studies are also needed about the effectiveness of statistics learning environments, learning environment design issues, the role of alignment between the various dimensions of statistics learning environments, new possibilities for teaching and learning in innovative designs, and opportunities in cutting-edge areas, such as model-based reasoning, visual representation to teach complex abstract concepts, learning in virtual worlds, net-based collaborative teams and communities, and big data (see Chaps. 1, 13, and 15 this volume).

The difficulty of demonstrating the effectiveness of the approach in Example III above raises profound methodological issues in researching learning environments. A traditional approach to research is one in which most variables are controlled as far as possible and the focus is on the unidimensional variable in question. The learning environment approach acknowledges a complex system or ecology in which such a methodology is not sustainable. Instead, a design research approach (Cobb et al., 2003; Gravemeijer & Cobb, 2013) is needed, where iterative design of the learning environment sensitizes the research team to the key mechanisms for learning within the design. Note, however, that in design research also, empirical data on what students gain from participating in the learning environment is indispensable. We recommend that more attention be given to methodological aspects of researching the design of learning environments.

Secondly, due to the proliferation of learning in online settings, there is an increase of designs for online learning communities such as MOOCs and virtual environments (e.g., Pratt, Griffiths, Jennings, & Schmoller, 2016; Wild, 2007). There is therefore a need to study designs for learning environments of the future (Jacobson & Reimann, 2010). We argue that taking a learning environment perspective can advance our understanding of the online learning arenas.

First steps in moving toward the learning environment perspective in the statistics education community are for researchers to consider the implications of this approach in their studies and for professional development to support teachers to consider how current curricula and materials align in the context of social, cultural, physical, psychological, and pedagogical components of a learning environment. Careful and steady change over a period of time, rather than a push for radical change, may lead to a successful implementation of a learning environment in the statistics education world, both among researchers and teachers.

References

Ainley, J., Gould, R., & Pratt, D. (2015). Learning to reason from samples: commentary from the perspectives of task design and the emergence of "big data". *Educational Studies in Mathematics, 88*(3), 405–412.

Ainley, J., & Pratt, D. (2014a). Chance re-encounters: 'computers in probability education' revisited. In D. Frischemeier, P. Fischer, R. Hochmuth, T. Wassong, & P. Bender (Eds.), *Using tools for learning mathematics and statistics* (pp. 165–177). New York: Springer.

Ainley, J., & Pratt, D. (2014b). Expressions of uncertainty when variation is partially determined. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, AZ, USA*. Voorburg, Netherlands: International Statistical Institute.

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal, 32*(1), 23–38.

Aridor, K., & Ben-Zvi, D. (2017). The co-emergence of aggregate and modelling reasoning. *Statistics Education Research Journal, 16*(2).

Arnold, P. (2014). *Statistical investigative questions: An enquiry into posing and answering investigative questions from existing data*. Unpublished doctoral thesis, University of Auckland, New Zealand.

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*(3), 243–260.

Baglin, J. (2013). Evaluating learning theory-based methods for improving the learning outcomes of introductory statistics courses (unpublished doctoral dissertation). RMIT University.

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: CD Beta Press.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenging of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht, Netherlands: Kluwer Academic Publishers.

Bakker, A., & Gravemeijer, K. (2006). An historical phenomenology of mean and median. *Educational Studies in Mathematics, 62*(2), 149–168.

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., et al. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *The Journal of the Learning Sciences, 7*(3-4), 271–311.

Barry, J. (2007). Acculturation. In J. Grusec & P. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 543–560). New York: Guilford Press.
Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1-2), 127–155.
Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics (CD-ROM)*. International Statistical Institute: Voorburg, Netherlands.
Ben-Zvi, D. (2007). Using wiki to promote collaborative learning in statistics education. *Technology Innovations in Statistics Education, 1*(1).
Ben-Zvi, D., & Aridor, K. (2016). Children's wonder how to wander between data and context. In D. Ben-Zvi & K. Makar (Eds.), *The teaching and learning of statistics: International perspectives* (pp. 25–36). Switzerland: Springer International Publishing.
Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM, 44*(7), 913–925.
Ben-Zvi, D., & Ben-Arush, T. (2014). Exploratory data analysis instrumented learning with TinkerPlots. In D. Frischemeier, P. Fischer, R. Hochmuth, T. Wassong, & P. Bender (Eds.), *Using tools for learning mathematics and statistics* (pp. 193–208). New York: Springer.
Ben-Zvi, D., Gil, E., & Apel, N. (2007). What is hidden beyond the data? Helping young students to reason and argue about some wider universe. In D. Pratt & J. Ainley (Eds.), *Reasoning about informal inferential statistical reasoning. Proceedings of the 5th International Research Forum on Statistical Reasoning, Thinking, And Literacy*. Warwick, UK: University of Warwick.
Bereiter, C. (1985). Towards a solution of the learning paradox. Review of Educational Research, 55(2), 201–226.
Bidgood, P., Hunt, N., & Jolliffe, F. (Eds.). (2010). *Assessment methods in statistical education: An international perspective*. West Sussex, UK: Wiley.
Biehler, R. (2003). Interrelated learning and working environments for supporting the use of computer tools in introductory classes. In International Statistical Institute (Ed.): *CD-ROM Proceedings of the IASE Satellite Conference on Statistics Education and the Internet, Max-Planck-Institute for Human Development, Berlin, 11-12 August 2003*, Voorburg, Netherlands.
Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–690). Springer.
Bielaczyc, K. (2006). Designing social infrastructure: Critical issues in creating learning environments with technology. *The Journal of the Learning Sciences, 15*(3), 301–329.
Bielaczyc, K., & Collins, A. (1999). Learning communities in classrooms: A reconceptualization of educational practice. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theories* (Vol. II, pp. 269–292). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
Boud, D., Keogh, R., & Walker, D. (1985). Promoting reflection in learning: A model. In D. Boud, R. Keogh, & D. Walker (Eds.), *Reflection: Turning experience into learning* (pp. 18–39). East Brunswick, NJ: Nichols.
Bransford, J., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229–272). Cambridge: The MIT Press.
Burgess, T. (2007). Investigating the nature of teacher knowledge needed and used in teaching statistics. Unpublished doctoral thesis, Massey University, New Zealand.
Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education - a Joint ICMI/IASE Study: The 18th ICMI Study* (pp. 57–69). Dordrecht: Springer.
Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education Journal, 1*(1).

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity, and compliance. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 151–192). New York: McGraw-Hill.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24*(2), 205–249.

Cobb, G. W. (1992). Report of the joint ASA/MAA committee on undergraduate statistics. In *In the American Statistical Association 1992 Proceedings of the Section on Statistical Education* (pp. 281–283). Alexandria, VA: American Statistical Association.

Cobb, G. W. (1993). Reconsidering statistics education: A national science foundation conference. *Journal of Statistics Education, 1*(1), 1–28.

Cobb, P. (1994a). Constructivism in mathematics and science education. *Educational Researcher, 23*(7), 4–4.

Cobb, P. (1994b). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23*(7), 13–20.

Cobb, P., & McClain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 375–396). Dordrecht, Netherlands: Kluwer Academic Publishers.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

Cobb, P., Gravemeijer, K., & Yackel, E. (2011). Introduction. In E. Yackel, K. Gravemeijer, & A. Sfard (Eds.), *A journey in mathematics education research, insights from the work of Paul Cobb* (pp. 109–115). Dordrecht: Springer.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*(1), 1–78.

Cognition and Technology Group at Vanderbilt. (1998). Designing environments to reveal, support, and expand our children's potentials. In S. A. Soraci & W. McIlvane (Eds.), *Perspectives on fundamental processes in intellectual functioning: A survey of research approaches* (Vol. 1, pp. 313–350). Stamford, CT: Ablex.

Collins, A. (1999). Design issues for learning environments. In S. Vosniadou, E. De Corte, R. Glaser, & H. Mandl (Eds.), *International perspectives on the psychological foundations of technology-based learning environments* (pp. 347–361). Hillsdale, NJ: Lawrence Erlbaum Associates.

Connor, D. (2002). *CensusAtSchool 2000: Creation to collation to classroom.* Paper presented at the 6th International Conference on Teaching Statistics (ICOTS-6) at Cap Town, South Africa. http://iase-web.org/documents/papers/icots6/2d1_conn.pdf.

Conway IV, B. M. (2015). A comparison of high school students' development of statistical reasoning outcomes in high and low statistical reasoning learning environments. Doctoral dissertation, Auburn University, USA.

Cuban, L. (2003). *Why is it so hard to get good schools?* New York: Teachers College, Columbia University.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. Jossey-Bass.

De Corte, E., Verschaffel, L., Entwistle, N., & van Merrienboer, J. (2003). *Powerful learning environments (Unravelling basic components and dimensions)*. UK: Emerald.

Edelson, D., & Reiser, B. (2006). Making authentic practices accessible to learning: Design challenges and strategies. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 335–354). New York, NY: Cambridge University Press.

Everson, M., Zieffler, A., & Garfield, J. (2008). Implementing new reform guidelines in teaching introductory college statistics courses. *Teaching Statistics, 30*(3), 66–70.

Franklin, C., & Garfield, J. (2006). The guidelines for assessment and instruction in statistics education (GAISE) project: Developing statistics education guidelines for pre K-12 and college courses. In G. F. Burrill (Ed.), *Thinking and reasoning about data and chance: Sixty-eighth NCTM yearbook* (pp. 345–375). Reston, VA: National Council of Teachers of Mathematics.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht, the Netherlands: Reidel.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht, the Netherlands: Reidel.
Gal, I., & Garfield, J. (Eds.). (1997). *The assessment challenge in statistics education*. Amsterdam: IOS Press.
Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372–396.
Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, the Netherlands: Springer.
Garfield, J., & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics, 31*(3), 72–77.
Garfield, J., Hogg, B., Schau, C., & Whittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education [Online], 10*(2).
Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. Mathematical Thinking and Learning, 13(1-2), 87–108.
Gil, E., & Ben-Zvi, D. (2014). Long term impact of the connections program on students' informal inferential reasoning. In K. Makar, B. de Sousa and R. Gould (Eds.), Sustainability in statistics education (Proceedings of the Ninth International Conference on Teaching Statistics, ICOTS9, July 2014). Voorburg, The Netherlands: International Association for Statistical Education and International Statistical Institute.
Gravemeijer, K. (2002a). Developing research, a course in elementary data analysis as an example. In F. Lin (Ed.), *Common sense in mathematics education* (pp. 43–68). Taipei, Taiwan: National Taiwan Normal University.
Gravemeijer, K. (2002b). *Emergent modeling as the basis for an instructional sequence on data analysis*. Paper presented at the 6th International Conference on Teaching Statistics (ICOTS-6) at Cape Town, South Africa.
Gravemeijer, K. (2004). Learning trajectories and local instruction theories as means of support for teachers in reform mathematics education. *Mathematical Thinking and Learning, 6*(2), 105–128.
Gravemeijer, K., & Cobb, P. (2013). Design research from the learning design perspective. In T. Plomp & N. Nieveen (Eds.), *Educational design research: An introduction (part A)* (pp. 72–113). Enschede: SLO (Netherlands Institute for curriculum development). Retrieved from http://downloads.slo.nl/Documenten/educational-design-research-part-a.pdf.
Heyd-Metzuyanim, E. (2013). The co-construction of learning difficulties in mathematics—teacher–student interactions and their role in the development of a disabled mathematical identity. *Educational Studies in Mathematics, 83*(3), 341–368.
Hickey, D. T., Kindfield, A. C., Horwitz, P., & Christie, M. A. (2003). Integrating curriculum, instruction, assessment and evaluation in a technology-supported genetics learning environment. *American Educational Research Journal, 40*(2), 495–538.
Hod, Y., & Ben-Zvi, D. (2015). Students negotiating and designing their collaborative learning norms: A group developmental perspective in learning communities. *Interactive Learning Environments, 23*(5), 578–594.
Jacobson, M., & Reimann, P. (Eds.). (2010). *Designs for learning environments of the future (international perspectives from the learning sciences)*. New York, NY: Springer Science+Business Media.
Jonassen, D., & Land, S. (Eds.). (2012). *Theoretical foundations of learning environments* (2nd ed.). New York, NY: Routledge.
Kali, Y., Tabak, I., Ben-Zvi, D., et al. (2015). Technology-enhanced learning communities on a continuum between ambient to designed: What can we learn by synthesizing multiple research perspectives? In O. Lindwall, P. Koschman, T. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the material conditions of learning: The computer supported collaborative learning conference* (Vol. II, pp. 615–622). Gothenburg, Sweden: The International Society of the Learning Sciences.
Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences, 21*(1), 45–83.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28–37.
Kohn, A. (1999). *The schools our children deserve: Moving beyond traditional classrooms and 'tougher standards'*. Boston: Houghton Mifflin.
Konold, C., & Miller, C. D. (2011). TinkerPlots: Dynamic Data Exploration (Version v2.0) [Computer software]. Available: https://www.tinkerplots.com/
Konold, C. & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. Journal for Research in Mathematics Education, 33(4), 259–289.
Lehrer, R. (2009). Designing to develop disciplinary knowledge: Modeling natural systems. *American Psychologist, 64*(8), 759–771.
Lehrer, R., & Pfaff, E. (2011). Designing a learning ecology to support the development of rational number: Blending motion and unit partitioning of length measures. In D. Y. Dai (Ed.), *Design research on learning and thinking in educational settings: Enhancing intellectual growth and functioning* (pp. 131–160). New York: Routledge.
MacGillivray, H., & Pereira-Mendoza, L. (2011). Teaching statistical thinking through investigative projects. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 109–120). New York: Springer.
Makar, K., & Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning, 13*(1-2), 1–4.
Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1-2), 152–173.
Makar, K., Bakker, A., & Ben-Zvi, D. (2015). Scaffolding norms of argumentation-based inquiry in a primary mathematics classroom. *ZDM, 47*(7), 1107–1120.
Manor, H., & Ben-Zvi, D. (2017). Students' emergent articulations of statistical models and modelling in making informal statistical inferences. *Statistics Education Research Journal, 16*(2).
Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123–137.
Moore, D. S. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association, 93*(444), 1253–1259.
Neumann, D. L., Hood, M., & Neumann, M. M. (2013). Using real-life data when teaching statistics: Student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal, 12*(2), 59–70.
Loveland, J. L. (2014). Traditional lecture versus an activity approach for teaching statistics: A comparison of outcomes (unpublished doctoral dissertation). Utah State University.
Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education (a Joint ICMI/IASE Study, the 18th ICMI Study)* (pp. 323–333). New York: Springer.
Pfannkuch, M., & Rubick, A. (2002). An exploration of students' statistical thinking with given data. *Statistics Education Research Journal, 1*(2), 4–21.
Piaget, J. (1978). *Success and understanding*. Cambridge, MA: Harvard University Press.
Pratt, D., Griffiths, G., Jennings, D. & Schmoller, S. (2016). *Tensions and compromises in the design of a MOOC for adult learners of mathematics and statistics*. Presented at the 13th International Congress on Mathematical Education, Hamburg.
Prediger, S., Gravemeijer, K., & Confrey, J. (2015). Design research with a focus on learning processes: An overview on achievements and challenges. *ZDM Mathematics Education, 47*(6), 877–891.
Reston, E., & Bersales, L. G. (2008). Reform efforts in training statistics teachers in the Philippines: Challenges and prospects. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*.
Rogoff, B. (1994). Developing understanding of the idea of communities of learners. *Mind, Culture, and Activity, 1*(4), 209–229.

Rogoff, B., Turkanis, C. G., & Bartlett, L. (2001). *Learning together children and adults in a school community*. London, UK: Oxford University Press.

Roseth, C. J., Garfield, J. B., & Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education, 16*(1), 1–15.

Savelsbergh, E. R., Prins, G. T., Rietbergen, C., Fechner, S., Vaessen, B. E., Draijer, J. M., et al. (2016). Effects of innovative science and mathematics teaching on student attitudes and achievement: A meta-analytic study. *Educational Research Review, 19*, 158–172.

Sawyer, R. K. (2014). Introduction. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 1–18). Cambridge, UK: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 397–417). New York: Cambridge University Press.

Sfard, A., & Cobb, P. (2014). Research in mathematics education: What can it teach us about human learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 545–564). Cambridge, UK: Cambridge University Press.

Sfard, A., & Linchevski, L. (1994). The gains and the pitfalls of reification - the case of algebra. *Educational Studies in Mathematics, 26*(2), 191–228.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education (part 1)* (pp. 205–238). Dordrecht, Netherlands: Kluwer.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114–145.

Slootmaeckers, K., Kerremans, B., & Adriaensen, J. (2014). Too afraid to learn: Attitudes towards statistics as a barrier to learning statistics and to acquiring quantitative skills. *Politics, 34*(2), 191–200.

Streefland, L. (1991). *Fractions in realistic mathematics education: A paradigm of developmental research*. Dordrecht: Kluwer Academic Publishers.

Treffers, A. (1987). *Three dimensions: A model of goal and theory description in mathematics education - the Wiskobas Project*. Dordrecht: Reidel.

van Hiele, P. M. (1986). Structure and insight: A theory of mathematics education. Orlando, FL: Academic Press.

Vosniadou, S., & Vamvakoussi, X. (2006). Examining mathematics learning from a conceptual change point of view: Implications for the design of learning environments. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present and future trends. Sixteen essays in honour of Erik De Corte* (pp. 55–70). Oxford, UK: Elsevier.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction, 11*(4), 381–419.

Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Watson, A., & Ohtani, M. (Eds.). (2015). *Task design in mathematics education: An ICMI study 22*. Cham: Springer.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.

Wild, C. (2007). Virtual environments and the acceleration of experiential learning. *International Statistical Review, 75*(3), 322–335.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*(4), 458–477.

Index

D. Ben-Zvi et al. (eds.), *International Handbook of Research in Statistics Education*, Springer International Handbooks of Education,
https://doi.org/10.1007/978-3-319-66195-7

C

I

J

K

L

Q

R