

ICME-13 Monographs

Gail Burrill
Dani Ben-Zvi *Editors*

Topics and Trends in Current Statistics Education Research

International Perspectives



 Springer

ICME-13 Monographs

Series editor

Gabriele Kaiser, Faculty of Education, Didactics of Mathematics, Universität Hamburg, Hamburg, Germany

Each volume in the series presents state-of-the art research on a particular topic in mathematics education and reflects the international debate as broadly as possible, while also incorporating insights into lesser-known areas of the discussion. Each volume is based on the discussions and presentations during the ICME-13 conference and includes the best papers from one of the ICME-13 Topical Study Groups, Discussion Groups or presentations from the thematic afternoon.

More information about this series at <http://www.springer.com/series/15585>

Gail Burrill · Dani Ben-Zvi
Editors

Topics and Trends in Current Statistics Education Research

International Perspectives

 Springer

Editors

Gail Burrill
Program in Mathematics Education
Michigan State University
East Lansing, MI, USA

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

ISSN 2520-8322

ISSN 2520-8330 (electronic)

ICME-13 Monographs

ISBN 978-3-030-03471-9

ISBN 978-3-030-03472-6 (eBook)

<https://doi.org/10.1007/978-3-030-03472-6>

Library of Congress Control Number: 2018960209

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Foreword

How opportune to have available in this one volume chapters highlighting rich international collaborative research in statistics education to examine the teaching and learning of statistics. It is especially relevant that much of this current research focuses on content, pedagogy, and learning at the school level and with teacher preparation. More often than not in the past and even currently, only the most able students have been encouraged to study statistics. However, we live in a constantly changing world that is structured around many different forms of data (traditional and non-traditional) that impact every individual daily. It is more urgent than ever that ALL students begin at a young age in school to develop data sense and statistical thinking. For this to happen, our school-level teachers and teacher educators must also develop data sense, conceptual understanding, and habits of mind for reasoning statistically. We need evolving research that investigates and understands how students and teachers develop statistical reasoning and how to provide accessibility to all students and teachers. For instance, researchers have previously dedicated much research to identifying student and teacher misconceptions. As seen in this volume, the research has now moved to trying different strategies and inventions that can help address these misconceptions. The hope is that these potential strategies will be further researched in different contexts and other countries.

The research presented in this volume also informs and provides credibility to the writing and evolution of educational policy documents establishing recommendations for optimal teaching and student learning. For example, in 2008, the American Statistical Association (ASA) document, *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE)*, utilized international research to recommend guidelines for teaching statistics at the school level—a groundbreaking document. This document has influenced national statistics standards in several countries. Ten years later, the recommendations are still relevant and essential but with the continual evolution of types of data, availability of data, technology available for analysis of data—the *Pre-K-12 GAISE* is being updated with current research informing the additional skills students need while still maintaining the spirit of the original document recommendations. In the ASA

document, *The Statistical Education of Teachers (SET)*, at the time of publication in 2015, few research-based guidelines were in place concerning what teachers need to know to effectively teach statistics. How exciting to read the chapters in this volume on the research focused on teacher preparation and teacher understanding enhancing the recommendations from *SET*.

Statistical education research is still an emerging vital field of study. This volume demonstrates that the research is valued worldwide, with collaborative efforts between different countries. I am grateful to the editors, Gail Burrill and Dani Ben-Zvi, and to the many authors and researchers, for this copious volume of imperative research that supports the writing and development of resources to evaluate student and teacher learning and that supports the professional development of teachers.

Athens, GA, USA

Christine Franklin
University of Georgia

Contents

Part I Student Understanding

| | |
|--|----|
| 1 Visualizing Chance: Tackling Conditional Probability Misconceptions | 3 |
| Stephanie Budgett and Maxine Pfannkuch | |
| 2 Students' Development of Measures | 27 |
| Christian Büscher | |
| 3 Students Reasoning About Variation in Risk Context | 51 |
| José Antonio Orta Amaro and Ernesto A. Sánchez | |
| 4 Students' Aggregate Reasoning with Covariation | 71 |
| Keren Aridor and Dani Ben-Zvi | |

Part II Teaching for Understanding

| | |
|--|-----|
| 5 Design for Reasoning with Uncertainty | 97 |
| Hana Manor Braham and Dani Ben-Zvi | |
| 6 Building Concept Images of Fundamental Ideas in Statistics: The Role of Technology | 123 |
| Gail Burrill | |
| 7 Informal Inferential Reasoning and the Social: Understanding Students' Informal Inferences Through an Inferentialist Epistemology | 153 |
| Maïke Schindler and Abdel Seidouvy | |
| 8 Posing Comparative Statistical Investigative Questions | 173 |
| Pip Arnold and Maxine Pfannkuch | |

| | |
|--|-----|
| Part III Teachers' Knowledge (Preservice and Inservice) | |
| 9 Pre-service Teachers and Informal Statistical Inference: Exploring Their Reasoning During a Growing Samples Activity | 199 |
| Arjen de Vetten, Judith Schoonenboom, Ronald Keijzer and Bert van Oers | |
| 10 Necessary Knowledge for Teaching Statistics: Example of the Concept of Variability | 225 |
| Sylvain Vermette and Annie Savard | |
| 11 Secondary Teachers' Learning: Measures of Variation | 245 |
| Susan A. Peters and Amy Stokes-Levine | |
| 12 Exploring Secondary Teacher Statistical Learning: Professional Learning in a Blended Format Statistics and Modeling Course | 265 |
| Sandra R. Madden | |
| 13 Statistical Reasoning When Comparing Groups with Software—Frameworks and Their Application to Qualitative Video Data | 283 |
| Daniel Frischemeier | |
| Part IV Teachers' Perspectives | |
| 14 Teachers' Perspectives About Statistical Reasoning: Opportunities and Challenges for Its Development | 309 |
| Helia Oliveira and Ana Henriques | |
| 15 A Study of Indonesian Pre-service English as a Foreign Language Teachers Values on Learning Statistics | 329 |
| Khairiani Idris and Kai-Lin Yang | |
| Part V Statistics Curriculum | |
| 16 A MOOC for Adult Learners of Mathematics and Statistics: Tensions and Compromises in Design | 351 |
| Dave Pratt, Graham Griffiths, David Jennings and Seb Schmoller | |
| 17 Critical Citizenship in Colombian Statistics Textbooks | 373 |
| Lucía Zapata-Cardona and Luis Miguel Marrugo Escobar | |
| 18 Critical Mathematics Education and Statistics Education: Possibilities for Transforming the School Mathematics Curriculum | 391 |
| Travis Weiland | |
| Index | 413 |

Contributors

José Antonio Orta Amaro Escuela Nacional para Maestras de Jardines de Niños, Mexico City, Mexico

Keren Aridor Faculty of Education, The University of Haifa, Haifa, Israel

Pip Arnold Karekare Education, Auckland, New Zealand

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

Hana Manor Braham Faculty of Education, The University of Haifa, Haifa, Israel

Stephanie Budgett Department of Statistics, The University of Auckland, Auckland, New Zealand

Gail Burrill Program in Mathematics Education, Michigan State University, East Lansing, MI, USA

Christian Büscher TU Dortmund University, Dortmund, Germany

Arjen de Vetten Section of Educational Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Daniel Frischemeier Paderborn University, Paderborn, Germany

Graham Griffiths UCL Institute of Education, London, UK

Ana Henriques Instituto de Educação, Universidade de Lisboa, Lisbon, Portugal

Khairiani Idris State Institute for Islamic Studies of Lhokseumawe, Aceh Province, Indonesia

David Jennings Independent Consultant and UCL Academic Visitor, London, UK

Ronald Keijzer Academy for Teacher Education, University of Applied Sciences iPabo, Amsterdam, The Netherlands

- Sandra R. Madden** University of Massachusetts Amherst, Amherst, MA, USA
- Luis Miguel Marrugo Escobar** Universidad de Antioquia, Medellín, Colombia
- Bert van Oers** Section of Educational Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
- Helia Oliveira** Instituto de Educação, Universidade de Lisboa, Lisbon, Portugal
- Susan A. Peters** University of Louisville, Louisville, KY, USA
- Maxine Pfannkuch** Department of Statistics, The University of Auckland, Auckland, New Zealand
- Dave Pratt** UCL Institute of Education, London, UK
- Ernesto A. Sánchez** Departamento de Matemática Educativa, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Mexico City, Mexico
- Annie Savard** McGill University, Montréal, Québec, Canada
- Maike Schindler** Faculty of Human Sciences, University of Cologne, Cologne, Germany
- Seb Schmoller** Independent Consultant and UCL Academic Visitor, London, UK
- Judith Schoonenboom** Department of Education, University of Vienna, Vienna, Austria
- Abdel Seidouy** School of Science and Technology, Örebro University, Örebro, Sweden
- Amy Stokes-Levine** University of Louisville, Louisville, KY, USA
- Sylvain Vermette** Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada
- Travis Weiland** Department of Mathematical Sciences, Appalachian State University, Boone, NC, USA
- Kai-Lin Yang** National Taiwan Normal University, Taipei, Taiwan
- Lucía Zapata-Cardona** Universidad de Antioquia, Medellín, Colombia

Introduction

Statistics is a general intellectual method that applies wherever data, variation, and chance appear. It is a fundamental method because data, variation, and chance are omnipresent in modern life

(Moore 1998, p. 134).

Background

The digital revolution (e.g., Hilbert and López 2011) of the last few decades coupled with the more recent data revolution (e.g., Kitchin 2014) has made statistical thinking and reasoning a necessity in today's world, a world awash with data that come in many different forms such as pictures, dynamic images, including innovative and interactive visualizations, and sounds as well as numbers and more traditional graphs. Analyzing those data is crucial to nearly every aspect of society including business, industry, social welfare, education, and government. The explosion of data has led to “big data” and the emerging need for increased attention to making sense of data as an integral part of many career options (Ben-Zvi 2017).

Being able to provide sound evidence-based arguments and critically evaluate data-based claims are important skills that all citizens should have. The study of statistics can provide students with tools, ideas, and dispositions to use in order to react intelligently to information in the world around them, rather than relying on subjective and often ill-informed judgments (Ben-Zvi and Makar 2016). To make this happen, the ability and inclination of learners, teachers, professionals, and citizens to understand, use, and communicate about statistics and probability need to be fundamental in educational programs, programs that are informed by research and evidence from practice that show promise for improving statistical teaching and learning (Garfield and Ben-Zvi 2007). Much can be learned by integrating results from such a variety of research and practice in statistics education. Such integration of theories, empirical evidence, and instructional methods can eventually help

students to develop their statistical thinking. These ongoing efforts to reform statistics instruction and content have the potential to both make the learning of statistics more engaging and prepare a generation of future citizens that deeply understand the rationale, perspective, and key ideas of statistics.

To that end, this volume is divided into five sections, each dealing with an aspect of teaching and learning statistics: student understanding, teaching for understanding, teachers' knowledge (preservice and inservice), teachers' perspectives, and curriculum.

ICME-13 Topic Study Group 15: Teaching and Learning of Statistics

This volume is a product of the Thirteenth International Congress on Mathematical Education (ICME-13) Topic Study Group 15 (TSG-15), Teaching and Learning Statistics. The members of TSG-15 came from 34 different countries and varied significantly by experience, background, and seniority. During the Congress, more than 60 presentations were divided into six themes related to key issues in statistics education research: core areas in statistics education; technology and the teaching of statistics; statistics education at the elementary level; statistics education at the secondary level; statistics education at the tertiary level; teachers' statistical knowledge and statistics education of preservice/in-service teachers; and future directions in statistics education.

The four meetings of TSG-15 were organized to create a sense of community among all presenters and participants, who shared a common desire and passion to improve statistics education by focusing on conceptual understanding rather than rote learning. The chapters in this volume are based on the 18 best papers presented in the four meetings.

Student Understanding

In Chap. 1, Budgett and Pfannkuch describe a study designed to address the difficulties students have in understanding conditional probabilities and Bayesian-type problems. Using a dynamic pachinkogram, a visual representation of the traditional probability tree, they explored six undergraduate probability students' reasoning processes as they interacted with the tool. Initial findings suggested that the ability to vary the branch widths of the pachinkogram may have the potential to support a more robust understanding of conditional probability.

In Chap. 2, Büscher invokes a theoretical framework for describing students' development of statistical concepts. A conceptualization of measure is introduced that links concept development to the development of measures, which consists of the three mathematizing activities of structuring phenomena, formalizing

communication, and creating evidence. An analysis of the results of a qualitative study in the framework of topic-specific design research reveals some impact of the context on students' situated concept development.

Orta and Sánchez explore students' reasoning concerning variation in Chap. 3. They analyzed the responses to two problems from a questionnaire administered to 87 ninth-grade students in which the students compared two data distributions and had to choose one they thought most advantageous for the situation. The authors propose three levels of reasoning related to how students interpret variation where decision making in the third level seems to be influenced by risk aversion or seeking.

In Chap. 4 Aridor and Ben-Zvi focus on understanding how students interpret and evaluate the relationship between two variables and the role of models in developing their reasoning. They used an illustrative case study to examine two students' emergent aggregate reasoning with covariation as the students explored the relations between two variables in a small real sample and constructed and improved a computerized statistical model of the predicted relations in the population using the software *TinkerPlotsTM*.

Teaching for Understanding

In Chap. 5, Manor Braham and Ben-Zvi focus on an “Integrated Modeling Approach” (IMA) that aspires to assist students to reason with the uncertainty involved in drawing conclusions from a single sample to a population. The chapter describes the design principles and insights arising from the implementation of one activity in the IMA learning trajectory in a case study of two students (age 12, grade 6). Implications for research and practice are also discussed.

The focus in Chap. 6 is on the use of dynamically linked documents based on *TI[©] Nspire* technology to provide students with opportunities to build coherent mental concept structures by taking meaningful statistical actions, identifying the consequences, and reflecting on those consequences, with appropriate instructional guidance. Burrill describes a collection of carefully sequenced documents based on research about student misconceptions and challenges in learning statistics. Initial analysis of data from preservice elementary teachers in an introductory statistics course highlights the students' progress in using the documents to cope with variability in a variety of contextual situations.

In Chap. 7, Schindler and Seidouvy present results from a study investigating the social nature of seventh-grade students' informal statistical inference (ISI) and informal inferential reasoning (IIR) in an experiment with paper helicopters. They describe how students draw inferences when working in a group and how the student inferences emerge socially with inferentialism used as a background theory. The results illustrate how students' informal inferences are socially negotiated in the group, how students' perceived norms influence IIR, and what roles statistical concepts play in students' IIR.

The focus in Chap. 8 by Arnold and Pfannkuch is on “good” statistical investigative questions that allow rich exploration of the data in hand, discovery, and thinking statistically. The work described in this chapter builds on earlier work related to the development of criteria for what makes a good statistical investigative question and a detailed two-way hierarchical classification framework for comparative statistical investigative questions that are posed. The authors’ focus is on the last of four research cycles, in which they explore responses from pre- and post-tests and discuss the level of comparative statistical investigative questions that students posed.

Teachers’ Knowledge (Preservice and Inservice)

In Chap. 9, de Vetten, Schoonenboom, Keijzer, and van Oers focus on the development of informal statistical inference (ISI) skills, not from the perspective of the students as in Chap. 7 but rather from the perspective of preservice teachers’ reasoning processes about ISI. Three classes of first-year elementary preservice teachers were asked to generalize from a sample to a population and to predict the graph of a larger sample during three rounds with increasing sample sizes. The analysis of the results revealed that most preservice teachers described only the data and showed limited understanding of how a sample can represent the population.

The focus of Chap. 10 is an exploration of teachers’ statistical knowledge in relation to the concept of variability. Vermette and Savard asked twelve high school mathematics teachers to respond to scenarios describing students’ strategies, solutions, and alternative conceptions when presented with a task in which variability was central to the interpretation. The authors analyzed the teachers’ comprehension and practices to gain insight into how to teach variability. The study found that students and high school teachers seem to share the same misconceptions related to the concept of variability.

Chapter 11 also considers teachers’ understanding of variation. Peters and Stokes-Levine describe results from a project to design and implement professional development for middle and high school mathematics teachers to investigate how dilemma, critical reflection, and rational discourse affect teachers’ understandings and reasoning about variation. Framed by transformative learning theory, the study highlights how teachers’ engagement with activities designed to prompt dilemma, consideration of multiple perspectives through multiple representations and rational discourse, and examination of premises underlying measures and procedures broadened the teachers’ perspectives on measures of variation.

Recognizing that many mathematics and science teachers in the USA have not benefitted from sufficient opportunity to learn statistics in a sense-making manner, in Chap. 12, Madden describes a study to support the statistical learning trajectory of in-service teachers. The study explores ways in which a course that blends face-to-face and virtual learning experiences impacted secondary in-service teachers’

technological pedagogical statistical knowledge (TPSK). Results suggest the course positively impacted participants' TPSK.

Chapter 13 introduces a framework for evaluating statistical reasoning and software skills when comparing groups. Frischemeier describes an application of this framework to qualitative data collected during a video study of four pairs of preservice teachers engaged in comparing distributions of data from different groups using *TinkerPlotsTM*. The results were used to evaluate the complex intertwined processes of the teachers' statistical reasoning and the use of software.

Teachers' Perspectives

In Chap. 14, Henriques and Oliveira investigated the perspectives of 11 mathematics teachers about the potential and the challenges of developing a learning environment targeting students' statistical reasoning in a developmental research project context. Findings show that the middle-grade teachers were able to distinguish key aspects that characterized the statistical reasoning in the tasks and ways the students used the software to explore the tasks, as well as recognizing that, as teachers, it is necessary to assume a new role that stands in contrast with traditional teaching practices.

In Chap. 15, Idris and Yang describe a phenomenographic approach to investigate the attitudes toward statistics of 38 Indonesian preservice teachers in an introductory college statistics course who were in an English as a Foreign Language (EFL) program. The authors identified three components of what was valued in learning statistics and related these to the components from task-value theory: intrinsic, attainment, and utility.

Statistics Curriculum

In Chap. 16, Pratt, Griffiths, Jennings, and Schmoller describe a project in the UK to develop a free open online course to offer motivated adults access to statistical ideas. The authors reflect on the tensions and compromises that emerged during the design of the course, in particular, the challenge of developing resources that will support heterogeneous students from unknown backgrounds, who may have already been failed by the conventional educational system and who will have no interactive tutor support within the online course.

Chapter 17 focuses on how the statistical component of fifth-grade mathematics textbooks in Colombia contributes to the development of critical citizenship using a socio-critical perspective. Zapata-Cardona and Marrugo-Escobar analyzed 261 tasks selected from seven mathematics textbooks. The results show that the contexts of the tasks were mostly hypothetical with very few tasks presented in real contexts. The tasks seemed to serve mainly as platforms to introduce

measurement calculations and application of statistical procedures, promoting procedural knowledge over reflective knowledge with little if any connection to a socio-critical perspective.

Chapter 18 discusses how ideas from critical mathematics education and statistics education intersect and could be used to transform the types of experiences that students have with both mathematics and statistics in the school mathematics curriculum. Weiland describes key ideas from the critical mathematics literature to provide a background from which to discuss what a critical statistics education could be. The chapter includes a discussion of some of the major barriers that need to be considered to make such a vision a reality and possible future directions for moving toward making a critical statistics education a reality.

Looking Forward: The International Perspective

The remarkable achievements of some countries in improving the teaching and learning of statistics have not yet arrived in all corners of the globe. Some countries still lack sufficient instructional resources, statistics curricular materials, effective professional development of preservice and in-service teachers, and educational technologies, foundations essential to carry on the reform movement in statistics education. Citizens in these countries are especially in need of becoming literate consumers of data, vital for improving their quality of life, monitoring and promoting social justice, economic growth, and the environment. They deserve, like any citizen of the world, to own the power of data literacy, to be able to add credibility to their claims and create and critically evaluate data-based evidence. Progress in the understandings of teaching and learning of statistical reasoning and thinking and the availability of high-quality technological tools for learning and teaching statistics should be shared with and by everyone, to help every country, region, school, and teacher worldwide to integrate and readily capitalize on these advances.

The studies in this book serve as a contribution in these directions. The chapters together create an arena for collaboration in synergetic cross-country research and development projects and nurture and encourage a sense of inclusiveness. It is essential to empirically test novel theoretical and practical ideas, successful in one context, in other countries and settings to learn more about their local and global affordances and constraints.

It is encouraging to see the diversity, creativity, richness, and novelty of the contributions across continents. It is a sound evidence for the growing numbers of enthusiastic and able scholars, the success of the statistics education community worldwide, and the increasing recognition that statistics education is receiving in the educational world, especially in the mathematics education community. While some countries are facing the enormous challenge of introducing statistics into the national curriculum for the first time, others are experimenting and evaluating with a second or a third wave of curricular reforms that already include strong

ingredients of data and chance in the school level. We embrace this diversity but urge all involved to increase international collaboration, sharing, and contribution, to the mutual benefit of all future citizens of the world.

References

- Ben-Zvi, D. (2017). Big data inquiry: Thinking with data. In R. Ferguson, S. Barzilai, D. Ben-Zvi, C.A. Chinn, C. Herodotou, Y. Hod, Y. Kali, A. Kukulska-Hulme, H. Kupermintz, P. McAndrew, B. Rienties, O. Sagy, E. Scanlon, M. Sharples, M. Weller & D. Whitelock (Eds.), *Innovating Pedagogies 2017: Open University Report 6* (pp. 32–36). Milton Keynes: The Open University.
- Ben-Zvi, D., & Makar, K. (Eds.) (2016). *The teaching and learning of statistics: International perspectives*. Springer International Publishing Switzerland.
- 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.
- Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332(6025), 60–65. <https://doi.org/10.1126/science.1200970>.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
- Moore, D. S. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association*, 93, 1253–1259.

Part I
Student Understanding

Chapter 1

Visualizing Chance: Tackling Conditional Probability Misconceptions



Stephanie Budgett and Maxine Pfannkuch

Abstract Probabilistic reasoning is essential for operating sensibly and optimally in the 21st century. However, research suggests that students have many difficulties in understanding conditional probabilities and that Bayesian-type problems are replete with misconceptions such as the base rate fallacy and confusion of the inverse. Using a dynamic pachinkogram, a visual representation of the traditional probability tree, we explore six undergraduate probability students' reasoning processes as they interact with this tool. Initial findings suggest that in simulating a screening situation, the ability to vary the branch widths of the pachinkogram may have the potential to convey the impact of the base rate. Furthermore, we conjecture that the representation afforded by the pachinkogram may help to clarify the distinction between probabilities with inverted conditions.

Keywords Bayesian-type problems · Conditional probability
Dynamic visualizations

1.1 Introduction

Our lives and the environments in which we live are pervaded by random events and chance phenomena. Events such as earthquakes, the global financial crisis, global warming and epidemics have resulted in many industries now paying attention to managing risk. The study of probability is a way of understanding the world from a non-deterministic perspective. Given the number of disciplines that require the application of probability concepts and understanding of probabilistic reasoning, the learning of probability is essential to prepare students for everyday life. Probability underpins the functioning of a modern economy and environment and, according to

S. Budgett (✉) · M. Pfannkuch
Department of Statistics, The University of Auckland, Auckland, New Zealand
e-mail: s.budgett@auckland.ac.nz

M. Pfannkuch
e-mail: m.pfannkuch@auckland.ac.nz

© Springer Nature Switzerland AG 2019
G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_1

Greer and Mukhopadhyay (2005, p. 308), is “a way of thinking [that] is supremely multidisciplinary.” Within a largely deterministic school curriculum, probability is the only subject area that exposes students to thinking about chance, learning to make decisions under uncertainty, and quantifying uncertainty. It is therefore indisputable that students need exposure to probabilistic thinking and reasoning. However, the current approach to teaching probability draws on the tradition of classical mathematical probability and may not be accompanied by a substantial understanding of the chance phenomena that the mathematics describes (Moore 1997). Often teaching approaches regress to a list of formulas and routine applications and, as Borovnick (2011, p. 81) observed, “probability is signified by a peculiar kind of thinking, which is not preserved in its mathematical conception.” Such an approach renders many probability ideas inaccessible to most students (Chernoff and Sriraman 2014).

A variety of strategies for improving people’s understanding of conditional probability have been investigated. Some of these approaches include displaying information in frequency format rather than in probability format (e.g. Gigerenzer and Hoffrage 1995) and providing accompanying static visualizations such as icon arrays and probability trees (e.g. Brase 2014; Garcia-Retamero and Hoffrage 2013). None of this research, however, has trialed a dynamic visualization approach. While the research approaches show improvement in people’s ability to deal with conditional probability scenarios with respect to the base rate fallacy and confusion of the inverse, advances in technology suggest that dynamic visualizations should be considered as another way to offer learners further insight into conditioning problems. We conjecture that if learners can vary input parameters such as the base rate and observe the resultant outputs, then they will experience and appreciate the effect of the base rate on the conditional outcomes at a deeper conceptual level. With regard to conditional probability, our research seeks to address a gap in the research knowledge base through exploring learners’ experiences as they interact with a dynamic probability tree, the pachinkogram, which visually represents proportions, distributions, randomness and variation.

The small exploratory study that forms the basis of this chapter has arisen as part of a larger research project which investigated the potential of exposing introductory university students to a modelling approach to probability involving learning strategies focused on dynamic visual imagery. The research question for this chapter is: How can a dynamic pachinkogram assist some students’ understanding of conditional probability through specifically raising their awareness of the base rate fallacy and confusion of the inverse?

1.2 Background Literature

An impoverished understanding of probabilistic information can lead to poor decision-making, with examples identified in fields such as medicine, management, law and intelligence analysis (Gigerenzer et al. 2007; Hoffrage et al. 2015; Mandel 2015; Nance and Morris 2005). Within the medical screening and diagnostic

field, Gigerenzer et al. (1998) document the tragic consequences of misinterpreting probabilistic reasoning.

Research in the field of cognitive science suggests that humans are innate Bayesians (Griffiths and Tenenbaum 2006; Pouget et al. 2013). Indeed it has been shown that infants as young as 12 months of age display Bayesian behavior in that they can integrate relevant information to form coherent judgments in unfamiliar situations (Téglás et al. 2011). However, literature within the field of probability education documents many misconceptions in peoples' thinking (Kahneman 2011). In particular, and despite people's intrinsic Bayesian behavior, Bayesian-type word problems present difficulties for many (Sirota et al. 2015), with the base rate fallacy and the confusion of the inverse misconception dominating people's reasoning (Villejoubert and Mandel 2002). The base rate fallacy arises most commonly when people neglect base rate information when judging probabilities (Bar-Hillel 1980; Kahneman and Tversky 1973), but can also arise when people give too much weight to base rate information (Teigen and Keren 2007). The confusion of the inverse misconception, where a conditional probability is confused with its inverse probability, has been documented by many and is often attributed to the existence of the base rate fallacy (Bar-Hillel 1980; Kahneman and Tversky 1973). However, other researchers have suggested that the base rate fallacy is an artefact of the confusion of the inverse misconception (Koehler 1996; Wolfe 1995).

Much effort has been directed into researching the pedagogical issues underlying Bayesian-type problems, resulting in several approaches designed to facilitate understanding. These approaches are now discussed.

1.2.1 Frequency Formats

Gigerenzer and colleagues noted that when probability information is presented as frequencies rather than in probability format, both experts and non-experts are less likely to succumb to the base rate fallacy and confusion of the inverse (Gigerenzer 2014). Despite the underlying mathematics being the same, Gigerenzer and Hoffrage (1995) demonstrated that when study participants were presented with a Bayesian problem where information was framed in a probability format, accuracy was 16%, and when the same information was framed in a frequency format, accuracy was 46%. It appears that, for most people, frequencies are computationally simpler to deal with than probabilities. Additionally, while base rate information is an integral component of frequency information, it is a less tangible component of probability information.

However, the format in which frequencies are presented is important. Although certain misconceptions may be alleviated when information is presented in frequency-format, Watson and Callingham (2014) noted that students have difficulty in employing proportional reasoning when interpreting frequency tables. In particular, they noted that teachers should "help students be flexible enough in their thinking

to consider proportions in each direction” (p. 279) highlighting the finding that most students in their study were unable to co-ordinate information in a 2×2 table.

1.2.2 Visualizations

Rapid advances in technology during the last few decades have led to the development of both static and dynamic visual representations which are now gaining momentum in the field of mathematics and statistics education. Clark and Paivio (1991) observed that visual images provide the opportunity for generation of mental images which can have a positive impact on learning. Furthermore, visual imagery can make visible concepts that were previously inaccessible within mathematical symbolic representations (Arnold et al. 2011; Konold and Kazak 2008; Pfannkuch et al. 2015). Arcavi (2003, p. 216) observed that cognitive technologies “might develop visual means to better ‘see’ mathematical concepts and ideas”. Because visualizations can bring to life properties that often remain intangible, they have the potential to support and augment students’ probabilistic reasoning. Examples of early visualizations tailored to the area of probability included blurred or degraded icons to represent varying levels of friendliness or hostility (Finger and Bisantz 2002) and the use of transparency, hue and opacity to represent uncertainty in weather predictions (Lefevre et al. 2005).

1.2.3 Visualizations for Bayesian-Type Problems

In the last two decades, researchers have investigated the performance of a variety of visual tools designed to improve people’s assessment of Bayesian-type situations. Sedlmeier and Gigerenzer (2001) demonstrated that the use of a frequency tree could improve respondents’ performance in a Bayesian-type task. More recently, Binder et al. (2015) investigated the effects of information format (probabilities or frequencies) and visualization (none, 2×2 table, tree diagram) on the performance of 259 German school students aged 16–18 years when presented with two Bayesian-type problems. They found that performance was optimized in the frequency/visualization condition, with no evidence of an effect of visualization type (2×2 table vs. tree diagram). Sloman et al. (2003) investigated the use of the Euler diagram as a visual representation of Bayesian-type probability information and found that 48% of the 25 respondents were able to provide an appropriate solution while Brase (2009) demonstrated, in three separate experiments, that participants randomized to the Euler or Venn diagram condition did not perform as well as participants randomized to an icon array condition. The icon array was developed by a risk communication specialist to assist those in the medical profession in particular (Paling 2003).

Several studies have demonstrated the effectiveness of the icon array as a means to visually represent information associated with Bayesian-type problems (e.g. Brase 2009, 2014; Zikmund-Fisher et al. 2014). Sedlmeier and Gigeren-

zer (2001), extending the frequency-based framework proposed by Gigerenzer and Hoffrage (1995), noted that the performance of people's interpretation of frequency-based information can be further enhanced by providing accompanying frequency-based static visualizations which have similarities to icon arrays. This finding was further reinforced in a study by Garcia-Retamero and Hoffrage (2013) who endorsed the use of visual aids such as frequency grids as an effective means of communicating quantitative medical information. The unit square has also been proposed as a visual tool which has the capacity to make transparent probability information (Bea 1995; Sturm and Eichler 2014). Indeed, in a study of prospective teachers, the unit square out-performed the tree diagram in terms of both procedural and conceptual knowledge (Böcherer-Linder et al. 2016). Additionally, and in light of the difficulties that many students experience when collating information from a 2×2 table (Watson and Callingham 2014), recent research has highlighted potential benefits of the eikosogram, a visual representation of a 2×2 table of information, in promoting proportional reasoning and flexibility in thinking (Pfannkuch and Budgett 2016a).

1.2.4 The Role of Technology

Technology's advances and ubiquitous access have provided an opportunity for students to visualize probability, or chance, through the creation of new representational infrastructure. Therefore it may be possible for students to gain access to previously inaccessible concepts (Sacristan et al. 2010). Furthermore, according to Shaughnessy (2007, p. 95), "technological tools are very important for helping students to transition from those naïve conceptions to richer more powerful understanding of statistical concepts". However, technology is not sufficient for conceptual growth. Both the teacher's and the student's articulation of how they make sense of, and explain in their own words, what they see and understand and thereby create meaning from the images, is critical (Makar and Confrey 2005). Reflective dialogue between teacher and students is paramount in developing conceptual reasoning (Bakker 2004).

1.2.5 Previous Related Work

In prior research, we interviewed seven practitioners who used stochastic modelling and probability in their work. The main purpose of the interviews was to understand the practitioners' view on the essential conceptual ideas required for probabilistic thinking. Together with a synthesis of related literature, we ascertained that randomness, distribution, conditioning and mathematics were the core foundational elements underpinning probabilistic thinking (Pfannkuch et al. 2016). The practitioners also identified conditioning and associated ideas as problematic areas for students. This is not surprising, and aligns with the fact that people have trouble making judgments in

Bayesian-type situations (Sirota et al. 2015). When asked about possible strategies to enhance students' understanding of probability, the key ideas that emerged were to: incentivize students to engage in understanding the ideas; use visual imagery; allow students to play around with chance-generating mechanisms; develop strategies to enable students to link across representations including extracting information from word problems; and use contexts that students can relate to.

Based on these key ideas, a six-principle framework was used to guide the design of a prototype dynamic visualization tool and associated task. The principles were to encourage students to (1) make conjectures, (2) test their conjectures against simulated data, (3) link representations, (4) perceive dynamic visual imagery, (5) relate to contexts, and (6) interact with chance-generating mechanisms. This framework was also used to analyze the data. These six principles are based on literature derived from a number of different sources. Principles (1) and (2) are key strategies which appear to promote active learning and student motivation (e.g. Lane and Peres 2006; Garfield et al. 2012), while principle (5) is also closely linked with student engagement (e.g. Neumann et al. 2013). The principle of linking representations was based upon the versatile thinking framework developed by Thomas (2008). The fourth principle aligns with recommendations that probability be taught from a modelling perspective and that simulation plays an important role in linking reality with probability models (Batanero et al. 2016). However, rather than being black-box abstractions, visually-based simulations allow students to experience random phenomena as they develop (Budgett et al. 2013). The final principle, that of having students interact with chance-generating mechanisms, is in response to Biehler's (1991) vision of having "more experiences with software where students can design random devices on the screen" (p. 189). Whether designing random devices, or modifying existing ones, students require many experiences across a range of contexts in order to painstakingly build "an abstract understanding of what to look for" (Cobb 2007, p. 339). The six-principle framework is discussed in more detail in Pfannkuch and Budgett (2016b).

1.3 Method

The aim of this exploratory study was to explore the potential of a dynamic visualization prototype tool and associated task to enhance introductory university students' understanding of probability. Given the problems people have in judging Bayesian-type situations, we were particularly interested in manifestations of the common misconceptions of confusion of the inverse and the base rate fallacy. To encourage the students to think aloud as they progressed through the task we used a two-person protocol in which the students discussed their thoughts and proposed actions. The research method is analogous to a pre-clinical trial in which a proposed intervention is investigated and modified in a laboratory setting prior to implementation in humans (Schoenfeld 2007). Occasionally the two authors would intervene for clarification purposes, or to progress the students if time became an issue.

1.3.1 Participants, Data Collection and Analysis

We randomly selected students ($n = 24$) who had successfully completed a first-year introductory probability paper ($n = 100$) until we had the consent of six students to participate. Ethics prevented us from conducting the study before or during the course. The introductory probability paper covered conditional probability using a traditional teaching approach. A pre- or co-requisite for enrolment into this introductory probability paper is successful completion of an introductory university general mathematics course which includes functions, linear equations and matrices, differential calculus of one and two variables, and integration of one variable. Therefore the participating students had an adequate grounding in mathematics and could be expected to cope with the material required to learn probability in the traditional manner. At the time of the study, one student planned to major in statistics, three planned to include statistics as a component in a double major or a conjoint degree, and two planned to major in disciplines other than mathematics and statistics but opted to take one or two statistics papers in their study programme. The pseudonyms of the pairs of participating students were: Brad and Ailsa, Harry and Hope, and Lorraine and Xavier. The task took approximately two hours to complete. The students were asked to think aloud and their dialogue was audio and video-taped. Screen captures, using Camtasia, recorded the students' interaction with the software tool. A post-interview was conducted after completion of the task in which students were asked to reflect on what they thought they had learned, and to suggest any improvements to the tool or accompanying task. Recordings made during the task sessions, the task-dialogue, and the post-interviews were transcribed. The task- dialogue transcription was qualitatively analyzed using a systematic process (cf. Braun and Clarke 2006) of (1) familiarizing oneself with the data, (2) searching for initial features in students' reasoning with regard to their conjectures and the testing of those conjectures against the simulated data and (3) identifying and reviewing critical and salient reasoning features that emerged about the base rate fallacy and confusion of the inverse in the new environment.

1.3.2 The Task and Tool

Students, in pairs, were presented with a task. Conjectures were sought from the students prior to introduction to the software tool and visualizations. Many researchers (e.g. Garfield et al. 2012; Konold and Kazak 2008) have suggested that seeking intuitions and conjectures from students helps them to engage with the task, to understand the situation presented, and provides motivation for exploring the problem. The context for the task is described in Fig. 1.1. The students were asked to read the information and to provide an intuitive answer to the question posed. The question in Fig. 1.1 was designed to address the confusion of the inverse misconception. Subsequent questions within the task involved the students thinking about the same

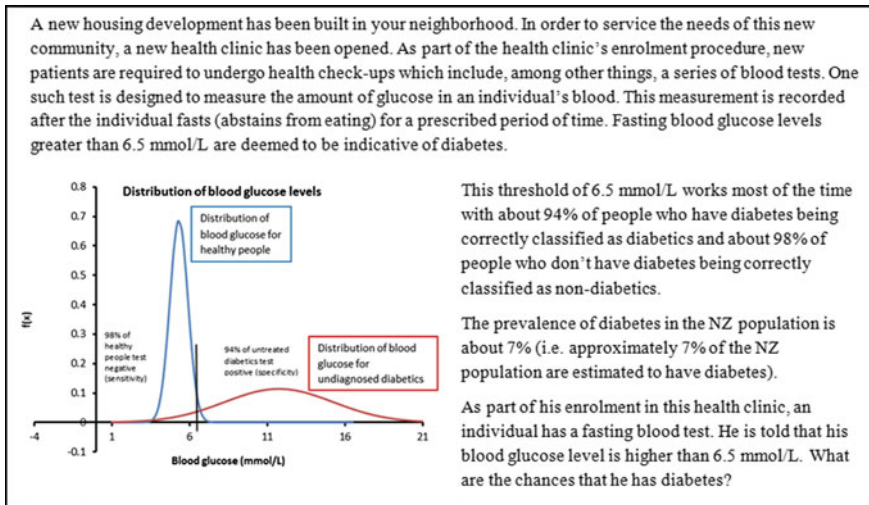


Fig. 1.1 Background information and question

situation, but considering different subgroups of the New Zealand population with varying diabetes prevalence which was based on real data (Coppell et al. 2013). For example, the prevalence for the New Zealand population as a whole is 7%. For New Zealand females aged between 25 and 34 the prevalence of diabetes is 3%, for New Zealand males aged between 65 and 74 the prevalence is 18.7%, and in Pacific Island people aged over 75 the prevalence is 56%. The rationale for providing students with the same problem for different subgroups of the NZ population was to raise awareness of the base rate fallacy. For some further details of the task, see [Appendix](#).

In accordance with the six-principle framework mentioned previously, the task was designed around a relatable context within a local setting. The students were first asked to make conjectures by answering each question intuitively before moving to the software tool. The software tool incorporated a variety of linked representations including symbolic representations and static and dynamic visual imagery. Having provided intuitive answers to questions within the task, the students interacted with the chance-generating mechanisms embedded within the software tool and simulated each scenario, thereby testing their conjectures against simulated data.

The students' first experience of a pachinkogram¹ is shown in Fig. 1.2 (www.stat.auckland.ac.nz/~vt). Note that the tool contains two additional components, an eikosogram and a graph, but we asked the students to focus on the pachinkogram to begin with. Figure 1.3a illustrates the pachinkogram set up to represent the situation described in Fig. 1.1 and Fig. 1.3b illustrates the pachinkogram set up to represent the same situation but with a base rate of diabetes of 56% which corresponded to the diabetes prevalence of Pacific Island people over the age of 75, one of the

¹The word pachinkogram originates from the Japanese pachinko machine which resembles a vertical pinball machine.

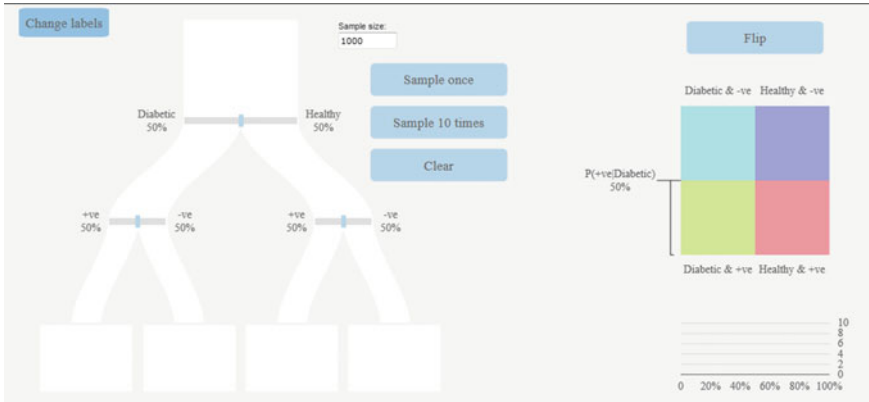


Fig. 1.2 Pachinkogram with default settings of equal probabilities on each branch (left), default eikosogram (top right) and empty empirical probability graph (bottom right)

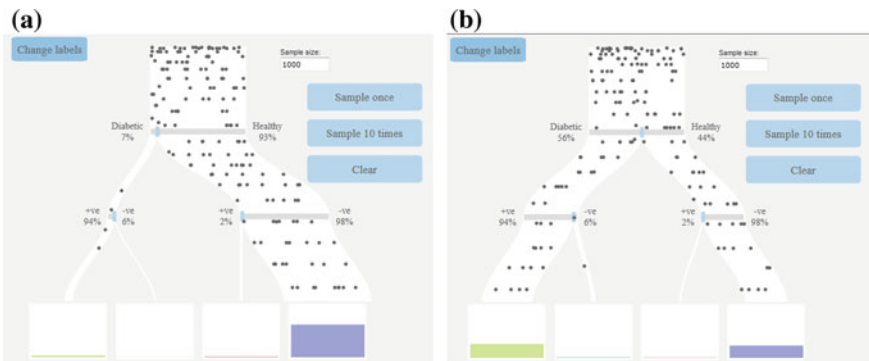


Fig. 1.3 Pachinkogram with **a** settings corresponding to the situation described in Fig. 1.1 and **b** settings corresponding to the situation described in Fig. 1.1 but with a base rate of 56%

subgroups presented to students in a subsequent question within the task. While the pachinkogram resembles the traditional probability tree, note that the branches of the pachinkogram are proportional in size to their respective probabilities. Probabilities on the pachinkogram branches can be changed by sliding the bars on the branches. Altering the input parameters of the pachinkogram in this way not only suggests to the user that the resultant output may change, but there is also a visible transformation in the appearance of the pachinkogram (e.g. compare Fig. 1.3a, b). For each simulation, dots corresponding to each member of the sample flow down the pachinkogram branches dynamically and end up in the ‘buckets’ at the bottom (see Fig. 1.3a, b).

One of the additional components of the tool, the eikosogram (see Fig. 1.2, top right), displays visual representations of simple, joint and conditional probabilities for each simulation. There is the facility to flip the condition and to toggle between

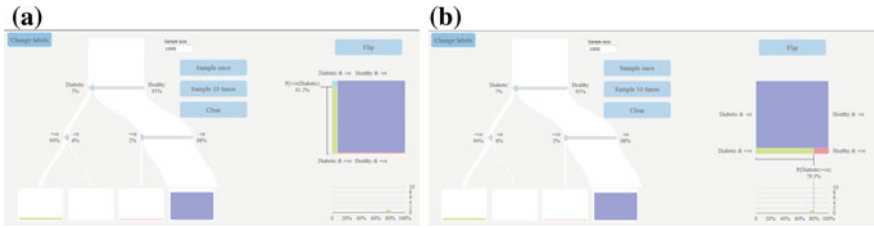


Fig. 1.4 Flipping the condition in the eikosogram (top right square) between **a** $P(+ve | Diabetic)$ and **b** $P(Diabetic | +ve)$

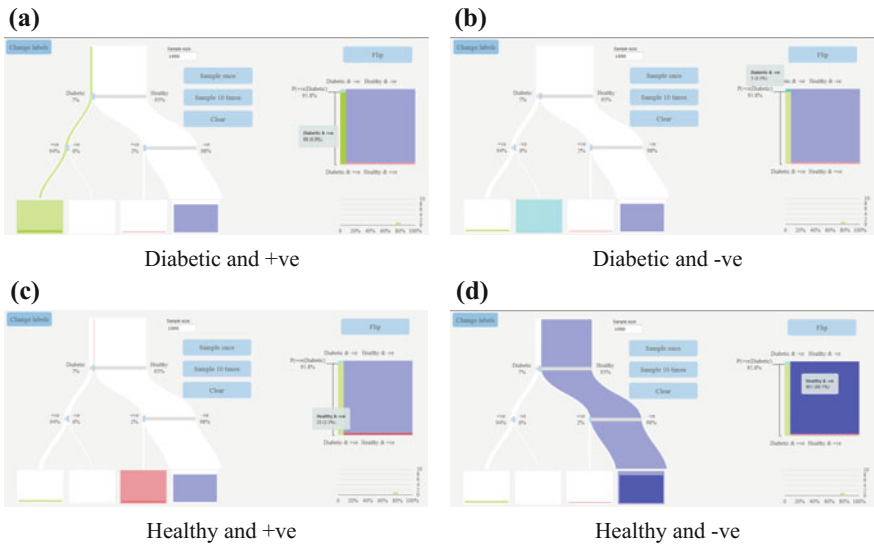


Fig. 1.5 Visual display of the dynamic linkage between pachinkogram pathways/buckets (left side) and the eikosogram joint probabilities (top right square)

$P(Diabetic | +ve)$ and $P(+ve | Diabetic)$ ² (see Fig. 1.4). This component is linked dynamically to the pachinkogram, highlighting the pathways and buckets representing the elements of a given joint probability (see Fig. 1.5).

The purpose of the task was to explore the students’ reasoning processes when interacting with simulated data displayed dynamically through a pachinkogram. A prominent feature of the task was to allow the students to make conjectures prior to using the tool, and then to test their conjectures with simulations. Of particular interest was their reaction to altering the pachinkogram parameters when changing

²Note +ve denotes a positive test result. Therefore $P(Diabetic | +ve)$ is the probability of having diabetes, given that a person has a positive test result and $P(+ve | Diabetic)$ is the probability of having a positive test result, given that a person is diabetic.

the base rate or the accuracy of the screening test (not explored in this paper), and the resulting effect on their initial conjectures.

1.4 Findings

Because the research question that guided the focus of this chapter is on the base rate fallacy and confusion of the inverse, and the impact of visualizations on these ideas, data relating to these issues has been drawn from the two-hour task sessions.

1.4.1 *Student Interaction with the Task and Tool*

Confusion of the inverse

Three of the six students, Xavier, Ailsa and Brad, appeared to demonstrate the confusion of the inverse misconception, that is they confused $P(\text{Diabetic} | +ve)$ with $P(+ve | \text{Diabetic})$ in their conjecture or intuitive answer to the question posed in Fig. 1.1, giving answers of around 95%. Only one student, Harry, provided a value (80%) that was close to the theoretical value of 78%. Hope gave an answer of 7% which suggested that, at this initial stage, she neglected to consider the fact that a positive test had been recorded, seeming only to focus on prevalence of diabetes in the population of interest and thereby demonstrating some level of confusion. Lorraine's answer was between 0.1 and 0.2 with the following explanation:

My intuitive answer used to be that the chances were pretty high but then I did stats and so now my intuitive answer would be that the chances are pretty low

It transpired that Lorraine had seen a similar question in a previous statistics course which was based on an example from an animation available on the Understanding Uncertainty website (Spiegelhalter n.d., <https://understandinguncertainty.org/screening>). This example involved screening for a particular attribute where the test had a reliability of 90% regardless of the presence or absence of the attribute, that is $P(+ve | \text{attribute present}) = 90\%$ and $P(-ve | \text{attribute absent}) = 90\%$, and the theoretical value of $P(\text{attribute present} | +ve)$ was 16%. Lorraine's memory of this example explained why she answered, incorrectly, that the chances of having diabetes given a positive test were 'pretty low'. Lorraine did not seem to have a sense of the magnitude of the difference between $P(\text{Diabetic} | +ve)$ and $P(+ve | \text{Diabetic})$ with a new base rate and a test which had an accuracy level that depended on an individual's attribute status. That is, she did not fully understand the impact of the base rate and the effect of the reliability of the test.

When the students were asked to modify the pachinkogram default settings in order to represent the situation presented to them, all of them appeared to gain more clarity on what the question was asking them. Even before the screening situation was

simulated, they were able to identify the characteristics of the ‘people’ who would end up in the buckets at the bottom of the pachinkogram and to comment on the resulting distribution. During Ailsa and Brad’s discussion of what they expected to see in the bucket distribution, Brad noted “I would expect very little in this one. It’s pretty unlikely [when referring to the bucket representing diabetics who test negative]”, while Ailsa stated that most of the people would end up in the negative and healthy bucket. Despite their initial conjectures demonstrating confusion of the inverse, their expectations for the distribution were in line with the result of the simulation. Xavier, who also exhibited confusion of the inverse in his original conjecture, mentioned that he expected more people in the leftmost bucket; that is, he expected to see more people who were diabetic being classified as diabetic. When Lorraine reminded him that only 7% of the population had diabetes, he admitted, “I didn’t think about the 7 percent.”

After viewing the simulation, they were asked if the bucket distribution corresponded with their expectations. Harry agreed, and proceeded to describe the characteristics of the people in the buckets stating “Most of the healthy people end up negative, and there’s a few poor people that got told they have diabetes when they don’t.” While gesturing to the bucket containing a small number of individuals who had diabetes but tested negative, Harry continued: “These are the most unlucky people because they have diabetes but they got told that they don’t.” He then completed his observation of the simulation by noting the people with diabetes who were correctly identified. Having performed one simulation, the students were able to focus on the buckets representing positive tests, and to note that a sizeable number were in fact non-diabetics who were incorrectly classified as diabetics. As Xavier noted, “I can see the visual of how it works”.

The students were then directed to the eikosogram in the top right of the screen and asked to consider the related probability (see Fig. 1.4a). They were specifically asked if the probability given on the screen was the answer to the original question in Fig. 1.1. When prompted to flip the eikosogram (see Fig. 1.4b), Ailsa identified the combined green and pink areas as representing all of the positive tests which seemed to allow her to recognise that the original question was asking for a probability conditional on testing positive, and not for a probability conditional on having diabetes. The following conversation between Brad and Ailsa demonstrates how they confronted their original misconception of confusion of the inverse and illustrates how Ailsa seemed to know what colour of buckets to focus on, and to link this to the same colours on the eikosogram.

- Brad:* Oh, we thought 95% intuitively
Interviewer: Has that answered the question?
Ailsa: It’s given diabetic that they test positive which is... ah...
Brad: Yeah, oh... that’s the other way around
Ailsa: That is given they are diabetic and test positive, whereas we were looking at it the other way around.

Xavier initially thought, incorrectly, that the probability from their simulation (similar to that in Fig. 1.4a) was the answer to the original question, although Lorraine was not convinced.

Interviewer: Is this the answer to question one?

Xavier: Yes, that is the answer to question one.

Lorraine: It is? Wait, wait, wait. I always get these round the wrong way. I always do it. No, isn't it the other way round? Isn't this one [question one] looking for diabetic given positive?

Xavier: Oh yeah, so [it] will be the opposite.

When Xavier and Lorraine flipped the eikosogram, they noted that the conditioning was changed with Xavier commenting, "That's the correct answer for question one. Before it was positive given diabetic and now it's diabetic given positive".

With three students displaying confusion of the inverse in their initial conjectures, and two students exhibiting other misconceptions, Harry was the only student who provided an approximately correct intuitive answer to the question in Fig. 1.1. When asked if the probability from their simulation (similar to that shown in Fig. 1.4a) was the answer to the original question, Harry stated: "No, because that is not the probability that it was positive, it's the opposite?". When asked to explain what he meant by 'opposite', he answered: "It should be like given the test was passed, what is the probability that she had diabetes".

Five of the six students' intuitive answers to the question in Fig. 1.1 demonstrated misunderstanding at some level. Instead of $P(\text{Diabetic} | +ve)$, Brad, Ailsa and Xavier all provided answers close to $P(+ve | \text{Diabetic})$ demonstrating confusion of the inverse. Hope's answer was $P(\text{Diabetic})$, suggesting that she failed to take account of the conditioning while Lorraine, who had seen a similar question before, provided the (incorrect) answer she recollected from that situation. However, after interacting with the pachinkogram and making links between the bucket distribution and the eikosogram, all of the students seemed to gain some clarity as to how and why $P(\text{Diabetic} | +ve)$ and $P(+ve | \text{Diabetic})$ were different.

Base rate fallacy

Although two pairs of students Ailsa and Brad, and Lorraine and Xavier, stated that the base rate was a necessary piece of information when asked for their intuitive answers to $P(\text{Diabetic} | +ve)$ in a variety of situations, they were not able to articulate *why*. Their initial intuitive answers to questions within the task also suggested that they were unaware of *how* the base rate might affect their answers. When Hope and Harry were asked if the base rate was a necessary piece of information in order to answer some of the task questions, they initially said no. Despite Harry's ability to recognise that $P(\text{Diabetic} | +ve)$ and $P(+ve | \text{Diabetic})$ were different probabilities, he was adamant that the base rate was not a required piece of information. However, having used the tool to answer some more of the questions, Harry and Hope quickly realised that a change in the base rate of diabetes in a population did have an effect on the numbers ending up in the buckets at the bottom of the pachinkogram, and hence on the value of $P(\text{Diabetic} | +ve)$. As Hope noted, "That is kind of funny, we said earlier

that we didn't think the percentage was actually relevant and now we are wanting to use it." In a subsequent section of the task, several seemingly identical questions were asked, although each had a different base rate. The following exchange illustrates Hope's realisation that an increase in the base rate from 7 to 18% will have an effect on the results of the simulation:

Interviewer: So have you got some expectation of what you are going to see in the buckets at the bottom?

Harry: This [pachinkogram] kind of makes it a little bit clearer

Hope: I think we are going to see a similar thing to last time, but we are going to have more in here [pointing to the leftmost bucket] and less in here [pointing to rightmost bucket] if that makes sense

Interviewer: Why are you going to have more in that left hand bucket?

Hope: Because you have increased the percentage of people who could have diabetes, so you are funnelling more down that side.

Hope's gestures and use of language suggests that it is the width of the pachinkogram branches that has provided her with further insight as to the impact of the base rate.

Similarly, through changing the base rate, all of the students now recognised that this would have an impact on the width of the branches of the pachinkogram and all were able to improve on their initial conjectures or intuitive answers. For example, when Xavier and Lorraine were asked, prior to running a simulation with a new base rate of 56% (see Fig. 1.3b), what distribution they expected to see in the buckets at the bottom of the pachinkogram, they responded:

Xavier: There will be more in this one [pointing to the leftmost bucket]

Lorraine: There will be quite a lot in the true positive one [pointing to the leftmost bucket]. These ones (the leftmost and rightmost buckets) will probably be relatively the same, I mean they will both be big ones. The other ones, not so much.

Xavier: But they [leftmost bucket] will be more than that one [rightmost bucket] because that one is bigger than that one.

Xavier is noting that there are now more diabetics than healthy people in the population (56% vs. 44%). It is not clear if he is also attending to the fact that the accuracy of the test depends on diabetes status. However, it appears that the width of the branches of the pachinkogram is contributing to his prediction.

When Ailsa and Brad altered the slider on the pachinkogram branches to reflect a new base rate of 56%, prior to running the simulation, Ailsa noted that she expected "more people correctly diagnosed as diabetic than people who are healthy and wrongly diagnosed". Her comment suggested that she now recognised the impact of the base rate and that in this new situation the majority of the positive tests would belong to people with diabetes. Although, initially, four of the six students believed that the base rate was a necessary piece of information that was required to answer the task questions similar to that shown in Fig. 1.1, none were able to articulate their reasoning. The remaining pair, Harry and Hope, were convinced that the base rate

was not required. After manipulating the pachinkogram branches to reflect new situations with varying base rates, the effect of the base rate appeared to become more transparent for all students.

1.4.2 *Student Reflections on the Task and Tool*

In the post-test interview, Lorraine and Xavier commented that they had more of an appreciation of the concept of conditional probability by being able to see both the dynamic simulation as the dots flowed through the pachinkogram and the widths of the pathways affecting the end result. Lorraine stated, “I guess I have more of an understanding on the probabilities of falling into each category than I did from [probability course] conditional probability.” Harry and Hope thought that the pachinkogram would be a useful tool when introducing conditional probability because, as Hope suggested, “it would be quite helpful to actually see we are conditioning on this [points to those testing positive] so we are ignoring the people who are going down that side and only focussing on these results.” Ailsa and Brad also discussed their new conditional probability understanding:

Ailsa: It’s a lot easier to see um how things change with changing the different thresholds, changing the diabetic versus healthy and then the success rates

Brad: Yeah, it’s like one small change can affect a lot of things

Ailsa: So it is a lot easier to understand the difference between which way around the conditional probability goes. And so how that relates, where those two numbers come from [$P(\text{Diabetic} | +ve)$ and $P(+ve | \text{Diabetic})$] and why they are different

Interviewer: It’s a misconception that those two probabilities are the same

Brad: No, they are definitely not the same.

Not only is Ailsa now aware of the importance of the conditioning, but she also appeared to understand why $P(\text{Diabetic} | +ve)$ and $P(+ve | \text{Diabetic})$ were different. This seems to be related to the visual aspect of the software tool, and in particular the colour connections that she made between the pachinkogram buckets and the eikosogram.

Lorraine reflected on her experience in the introductory probability course she had completed. She recalled being instructed to tackle problems such as the one described by constructing a 2×2 table of frequency information. However, when it came to answering a question under exam conditions, she had forgotten how to construct the table, stating, “I had forgotten how we did it...a whole lot of equations, one after the other, numbers, numbers, numbers.”

Although Ailsa and Brad decided that the base rate was a necessary piece of information, they were unable to articulate *why* this was the case and were initially unaware of the effect of a change in the base rate. However, reflecting on their experience with the software tool they remarked on the fact that a change in base

rate corresponded to a *visual* difference in the pachinkogram. Brad commented that “Changing (increasing) the percentage of the population that had diabetes, more of the little dots would go over to having diabetes, so there is a higher proportion testing positive and that had diabetes.” Ailsa thought that while a probability tree might have numerical values alongside the branches, “the buckets down the bottom being an area is a lot easier to see”, with Brad saying “I like how the things end up there. It helps.” Harry also indicated that the physical process of adjusting the pachinkogram branches to represent a given situation helped his reasoning, “Yeah, even just moving the slider across gives more intuition I think.”

The conversations above, in conjunction with accompanying gestures, illustrate that these students were attending to the width of the pachinkogram branches and anticipating the resulting effects. All six students specifically mentioned the visual aspect of the tool helping their understanding with Lorraine commenting, “A visual, instead of just one learning tool you’ve got two. Visual might be better for some, written might be better for some but if you’ve got both then you’re covering both of those, and both is surely better than just one.”

When compared to probability problems previously encountered by these students, the context of the problem seemed to incentivise them. Hope commented: “I like the background story... I like to know why I am doing it and I am not just computing random numbers for the sake of computing random numbers. I don’t like that. So a little bit of background definitely helps me get into it and think about things.” All of the students could identify with the diabetes screening context, and were familiar with the notion of varying diabetes prevalence rates according to gender and ethnicity. Although Harry and Hope did not initially think that the base rate was necessary in order to answer the first question, and Harry noted that he “felt quite sure about that at the time”, it only took one interaction with the pachinkogram for them to realise that if the screening test accuracy remained the same, a change in base rate had to have an effect. In Hope’s words: “The more you do it and the more you see how things affect it and change it and even just little changes as opposed to big changes you kind of get better at predicting it. So I kind of have more of a feel for where the numbers should lie just because I have seen a couple of examples now.”

1.5 Discussion

According to the cognitive theory of multimedia learning (Mayer 2009), advances in both graphical and information technology have paved the way for researchers to explore the potential of dynamic visualizations to facilitate access to a variety of concepts. Designers of such visualizations must be mindful of how visual information is processed (Ware 2008) and be aware of the fact that the format of a graphical display significantly affects how that display is understood (Mayer 2010).

We were interested in exploring the role of a dynamic visualization to support students’ probabilistic reasoning, with particular focus on Bayesian-type problems. Furthermore, since we believed that student engagement was an important concept

for the learning of probability, we designed a task and tool around a relatable context and incentivized students to make conjectures prior to interacting with technological tools. The pachinkogram and associated task appeared to assist this small group of students to recognize the base rate fallacy and confusion of the inverse.

Using the six-principle framework we now discuss briefly how each component seemed to enhance these students' probabilistic reasoning. A major feature that we observed in this study was student engagement. We attribute this to two components: the contextually relatable medical screening situation embedded within a local setting; and having the students put a stake in the ground through making conjectures throughout the task. The principles of *making* and *testing conjectures*, in conjunction with *providing a relatable context*, appeared to incentivize students not only to engage in the problem, but also to understand why their original conjectures might have been based on one or more misconceptions. The ability of simulations to assist students to understand why their initial conjectures might be incorrect has also been noted in other research (e.g., Konold and Kazak 2008). The *visual* representation afforded by the pachinkogram in Fig. 1.3a, b, notably the ability to change the widths of the branches to reflect the probabilities, seemed to be a powerful way to convey the impact of the base rate. This seemed to be further facilitated by the *linking of* the numerical *representations* of 7% and 93% with the widths of the pachinkogram branches (Fig. 1.3a).

We conjecture that the dynamic connections made between the pachinkogram buckets and the eikosogram (Fig. 1.5), including the colour connections, clarified the links between the properties of the 'people' in the buckets and the respective simple, joint and conditional probabilities represented in the eikosogram and that the flip facility (Fig. 1.4) aided the distinction between probabilities with inverted conditions. Furthermore, when simulating the screening situation, the utility of *visualizing* 'people' flowing through the pachinkogram to one of the four buckets at the base seemed to help in clarifying the probability asked for in the original question and differentiating it from its inverse (cf. Arcavi 2003). The simulation itself provided a frequency-type view of the screening situation, with each dot representing a person, and seemed to help these students clarify the required conditional probabilities, in line with the findings of Gigerenzer and others (Gigerenzer 2014).

In accordance with recent research (e.g. Garcia-Retamero and Hoffrage 2013), the visual aspect of the pachinkogram appeared to provide additional support for student understanding. Additionally, *linking* the branches and the buckets of the pachinkogram to symbolic *representations* promotes a flexible way of thinking which appeared to consolidate the students' understanding. The students' *interaction with a chance-generating mechanism*, that is the pachinkogram, reinforced the notion that the distribution of the number of people in each of the four buckets would vary from simulation to simulation, despite none of the pachinkogram parameters changing.

Overall we conjecture that these six students were beginning to develop some probabilistic intuition, including notions of randomness, conditioning and distribution which Pfannkuch et al. (2016) identify as core components associated with the development of probabilistic reasoning.

The four students who initially stated that the base rate was an essential piece of information displayed some form of confusion when providing an intuitive answer for $P(\text{Diabetic} \mid +ve)$. Therefore, although aware of the importance of the base rate, they exhibited the base rate fallacy by underestimating its effect. This resulted in three of the students succumbing to the confusion of the inverse misconception by mixing up $P(\text{Diabetic} \mid +ve)$ and $P(+ve \mid \text{Diabetic})$. This outcome is line with the findings of researchers who suggest that the confusion of the inverse misconception originates from the base rate fallacy (e.g. Bar-Hillel 1980; Kahneman and Tversky 1973). However, Harry's preliminary belief that the base rate was not required seems to be in conflict with the fact that he was able to distinguish $P(\text{Diabetic} \mid +ve)$ from $P(+ve \mid \text{Diabetic})$. Research has shown that study participants who were trained to distinguish $P(A \mid B)$ from $P(B \mid A)$ were less likely to demonstrate base rate neglect compared with a control group (Wolfe 1995). One possibility is that Harry was attending to the base rate of 7%, albeit unconsciously, or the other possibility is that Harry made a lucky estimate. Lorraine, on the other hand, could distinguish between the two conditions and believed that the base rate was involved, yet could not give a reasonable estimate.

Because this exploratory study only involved six students we cannot draw any conclusions about the link between the confusion of the inverse misconception and the base rate fallacy. All we can conclude, very tentatively, is that the pachinkogram seemed to assist these students to understand, at a visual level, *why* the base rate affected the bucket distributions, and therefore affected the resulting probabilities. The use of colour connections and dynamic linking between the buckets and the eikosogram appeared to help them to distinguish between probabilities with inverted conditions. Furthermore, the widths of the pachinkogram branches helped them to recognize *why* the probabilities with inverted conditions were different. Also, the participating students had already completed an introductory probability course and interacted with the task and tools in a semi-structured environment and hence further research in different settings and with students of varying ages and abilities is warranted.

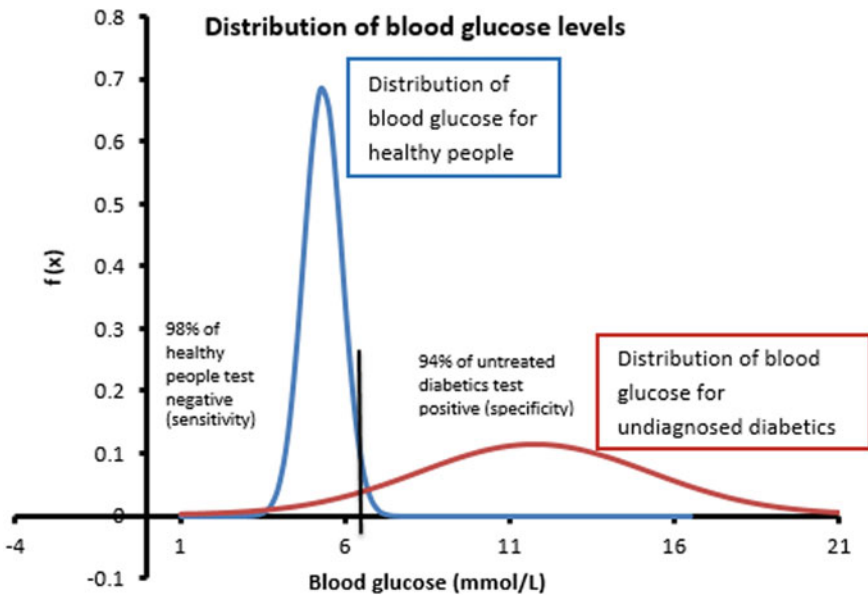
Random events and chance phenomena permeate our lives and environments, with many of the decisions made by citizens of the 21st century involving some level of probabilistic reasoning. Often poor judgments occur when people are required to assess or estimate Bayesian-type probabilities (Sirota et al. 2015), an area replete with misconceptions (Kahneman 2011). The traditional mathematical approach to teaching probability has resulted in many students unable to gain access to fundamental probability concepts (Chernoff and Sriraman 2014). Using the pachinkogram, the dynamic visualization tool described in this chapter, may raise awareness of common misconceptions such as the base rate fallacy and confusion of the inverse (Villejoubert and Mandel 2002) and instill in students a better understanding of core foundational elements underpinning probabilistic thinking (Pfannkuch et al. 2016).

Acknowledgements This work is supported by a grant from the Teaching and Learning Research Initiative (<http://www.tlri.org.nz/>).

Appendix: Some of the Diabetes Task Questions

A new housing development has been built in your neighbourhood. In order to service the needs of this new community, a new health clinic has opened. As part of the health clinic’s enrolment procedure, new patients are required to undergo health check-ups which include, among other things, a series of blood tests. One such test is designed to measure the amount of glucose in an individual’s blood. This measurement is recorded after the individual fasts (abstains from eating) for a prescribed period of time. Fasting blood glucose levels in excess of 6.5 mmol/L are deemed to be indicative of diabetes. This threshold of 6.5 mmol/L works most of the time with about 94% of people who have diabetes being correctly classified as diabetics and about 98% of those not having diabetes being correctly classified as non-diabetics.

The prevalence of diabetes in the NZ population is about 7% (i.e. approximately 7% of the NZ population are estimated to have diabetes).



Graph above adapted from Pfannkuch et al. (2002, p. 28)

Question 1

- As part of enrolment in this health clinic, an individual has a fasting blood test. He/she is told that his/her blood glucose level is higher than 6.5 mmol/L. What are the chances that he/she has diabetes? Provide an intuitive answer.
- Now use the software tool to answer Question 1 (a).
- Reflecting on your answer to (b), how does this compare with your answer to (a)?

Question 2

- (a) As part of his enrolment in this health clinic, a male aged between 65 and 74 has a fasting blood test.³ He is told that his blood glucose level is higher than 6.5 mmol/L. What are the chances that he has diabetes? Provide an intuitive answer.
- (b) Now use the software tool to answer Question 2 (a).
- (c) Reflecting on your answer to (b), how does this compare with your answer to (a)?

Question 3

- (a) As part of their enrolment in this health clinic, a person of Pacific ethnicity and aged over 75 has a fasting blood test. He/she is told that his/her blood glucose level is higher than 6.5 mmol/L. What are the chances that he/she has diabetes? Provide an intuitive answer.
- (b) Now use the software tool to answer Question 3 (a).
- (c) Reflecting on your answer to (b), how does this compare with your answer to (a)?

References

- Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics*, 52, 215–241.
- Arnold, P., Pfannkuch, M., Wild, C., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands-on to "movies". *Journal of Statistics Education*, 19(2), 1–32. Retrieved from <http://www.amstat.org/publications/jse/v19n2/pfannkuch.pdf>.
- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64–83.
- Bar-Hillel, M. (1980). The base rate fallacy in probability judgments. *Acta Psychologica*, 44, 211–233.
- Batanero, C., Chernoff, E., Engel, J., Lee, H., & Sánchez, E. (2016). Research on teaching and learning probability. In *Proceedings of Topic Study Group 14 at the 13th International Conference on Mathematics Education (ICME)*, Hamburg, Germany (pp. 1–33). https://doi.org/10.1007/978-3-319-31625-3_1.
- Bea, W. (1995). *Stochastisches denken [Statistical reasoning]*. Frankfurt am Main, Germany: Peter Lang.
- Biehler, R. (1991). Computers in probability education. In R. Kapadia & M. Borovnick (Eds.), *Chance encounters: Probability in education* (pp. 169–211). Boston, MA: Kluwer Academic Publishers.
- Binder, K., Krauss, S., & Bruckmaier, G. (2015). Effects of visualizing statistical information—An empirical study on tree diagrams and 2×2 tables. *Frontiers in Psychology*, 6(1186). <https://doi.org/10.3389/fpsyg.2015.01186>.
- Böcherer-Linder, K., Eichler, A., & Vogel, M. (2016). The impact of visualization on understanding conditional probabilities. In *Proceedings of the 13th International Congress on Mathemati-*

³Students were provided with a table of data from Coppell et al. (2013) from which they could determine the base rate for the subgroup of the population referred to in the questions.

- cal Education, Hamburg* (pp. 1–4). Retrieved from http://iase-web.org/documents/papers/icme13/ICME13_S1_Boechererlinder.pdf.
- Borovnick, M. (2011). Strengthening the role of probability within statistics curricula. 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. 71–83). New York, NY: Springer.
- Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology*, 23, 369–381. <https://doi.org/10.1002/acp.1460>.
- Brase, G. L. (2014). The power of representation and interpretation: Doubling statistical reasoning performance with icons and frequentist interpretation of ambiguous numbers. *Journal of Cognitive Psychology*, 26(1), 81–97. <https://doi.org/10.1080/20445911.2013.861840>.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- 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>.
- Chernoff, E. J., & Sriraman, B. (Eds.). (2014). *Probabilistic thinking: Presenting plural perspectives*. Dordrecht, The Netherlands: Springer. <https://doi.org/10.1007/978-94-007-7155-0>.
- Clark, J., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210.
- Cobb, G. (2007). One possible frame for thinking about experiential learning. *International Statistical Review*, 75(3), 336–347.
- Coppell, K. J., Mann, J. I., Williams, S. M., Jo, E., Drury, P. L., Miller, J., et al. (2013). Prevalence of diagnosed and undiagnosed diabetes and prediabetes in New Zealand: Findings from the 2008:2009 Adult Nutrition Survey. *The New Zealand Medical Journal*, 126(1370), 23–43.
- Finger, R., & Bisantz, A. M. (2002). Utilizing graphical formats to convey uncertainty in a decision-making task. *Theoretical Issues in Ergonomics Science*, 3(1), 1–25. <https://doi.org/10.1080/14639220110110324>.
- Garcia-Retamero, R., & Hoffrage, U. (2013). Visual representation of statistical information improves diagnostic inferences in doctors and patients. *Social Science and Medicine*, 83, 27–33.
- Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM—International Journal on Mathematics Education*, 44(7), 883–898.
- Gigerenzer, G. (2014). *Risk savvy: How to make good decisions*. New York, NY: Viking.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Bulletin*, 102, 684–704.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8, 53–96.
- Gigerenzer, G., Hoffrage, U., & Ehert, A. (1998). AIDS counseling for low-risk clients. *AIDS Care*, 10, 197–211. <https://doi.org/10.1080/09540129850124451>.
- Greer, B., & Mukhopadhyay, S. (2005). Teaching and learning the mathematization of uncertainty: Historical, cultural, social and political contexts. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 297–324). New York, NY: Kluwer/Springer Academic Publishers.
- Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, 17, 767–773. <https://doi.org/10.1111/j.1467-9280.2006.01780.x>.
- Hoffrage, U., Hafenbrädl, S., & Bouquet, C. (2015). Natural frequencies facilitate diagnostic inferences of managers. *Frontiers in Psychology*, 6(642), 1–11. <https://doi.org/10.3389/fpsyg.2015.00642>.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Allen Lane.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.

- Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences*, *19*, 1–17. <https://doi.org/10.1017/S0140525X00041157>.
- Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, *2*(1). Retrieved from <http://escholarship.org/uc/item/38p7c94v>.
- Lane, D. M., & Peres, S. C. (2006). Interactive simulations in the teaching of statistics: Promise and pitfalls. In B. Phillips (Ed.), *Proceedings of the Seventh International Conference on Teaching Statistics, Cape Town, South Africa*. Voorburg, The Netherlands: International Statistical Institute.
- Lefevre, R. J., Pfautz, J., & Jones, K. (2005). Weather forecast uncertainty management and display. In *Proceedings of the 21st International Conference on Interactive Information Processing Systems (UPS) for Meteorology, Oceanography, and Hydrology, San Diego, CA*. Retrieved from <https://ams.confex.com/ams/pdfpapers/82400.pdf>.
- Makar, K., & Confrey, J. (2005). “Variation-Talk”: Articulating meaning in statistics. *Statistics Education Research Journal*, *4*(1), 27–54.
- Mandel, D. R. (2015). Instruction in information structuring improves Bayesian judgment in intelligence analysis. *Frontiers in Psychology*, *6*(387), 1–12. <https://doi.org/10.3389/fpsyg.2015.00387>.
- Mayer, R. E. (2009). *Multimedia learning*. New York: Cambridge University Press.
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning graphics. *Learning and Instruction*, *20*, 167–171. <https://doi.org/10.1016/j.learninstruc.2009.02.012>.
- Moore, D. (1997). Probability and statistics in the core curriculum. In J. Dossey (Ed.), *Confronting the core curriculum* (pp. 93–98). Washington, DC: Mathematical Association of America.
- Nance, D. A., & Morris, S. B. (2005). Juror understanding of DNA evidence: An empirical assessment of presentation formats for trace evidence with a relatively small random-match probability. *Journal of Legal Studies*, *34*, 395–444. <https://doi.org/10.1086/428020>.
- 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. Retrieved from [https://iase-web.org/documents/SERJ/SERJ12\(2\)_Neumann.pdf](https://iase-web.org/documents/SERJ/SERJ12(2)_Neumann.pdf).
- Paling, J. (2003). Strategies to help patients understand risks. *British Medical Journal*, *327*, 745–748. <https://doi.org/10.1136/bmj.327.7417.745>.
- Pfannkuch, M., & Budgett, S. (2016a). Reasoning from an Eikosogram: An exploratory study. *International Journal of Research in Undergraduate Mathematics Education*, 1–28. <https://doi.org/10.1007/s40753-016-0043-0>.
- Pfannkuch, M., & Budgett, S. (2016b). Markov processes: Exploring the use of dynamic visualizations to enhance student understanding. *Journal of Statistics Education*, *24*(2), 63–73. <https://doi.org/10.1080/10691898.2016.1207404>.
- 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. 95–127). Minneapolis, MN: Catalyst Press.
- Pfannkuch, M., Budgett, S., Fewster, R., Fitch, M., Pattenwise, S., Wild, C., et al. (2016). Probability modelling and thinking: What can we learn from practice? *Statistics Education Research Journal*, 11–37. Retrieved from [http://iase-web.org/documents/SERJ/SERJ15\(2\)_Pfannkuch.pdf](http://iase-web.org/documents/SERJ/SERJ15(2)_Pfannkuch.pdf).
- Pfannkuch, M., Seber, G. A., & Wild, C. J. (2002). Probability with less pain. *Teaching Statistics*, *24*(1), 24–30.
- Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nature Neuroscience*, *16*, 1170–1178. <https://doi.org/10.1038/nn.3495>.
- Sacristan, A., Calder, N., Rojano, T., Santos-Trigo, M., Friedlander, A., & Meissner, H. (2010). The influence and shaping of digital technologies on the learning—and learning trajectories—of mathematical concepts. In C. Hoyles, & J. Lagrange (Eds.), *Mathematics education and technology—Rethinking the terrain: The 17th ICMI Study* (pp. 179–226). New York, NY: Springer.
- Schoenfeld, A. (2007). Method. In F. Lester (Ed.), *Second handbook of research on the teaching and learning of mathematics* (pp. 96–107). Charlotte, NC: Information Age Publishers.

- Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, 3, 380–400. <https://doi.org/10.1037//0096-3445.130.3.380>.
- 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.
- Sirota, M., Vallée-Tourangeau, G., Vallée-Tourangeau, F., & Juanchich, M. (2015). On Bayesian problem-solving: Helping Bayesians solve simple Bayesian word problems. *Frontiers in Psychology*, 6(1141), 1–4. <https://doi.org/10.3389/fpsyg.2015.01141>.
- Sloman, S. A., Over, D. E., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, 91, 296–309.
- Spiegelhalter, D. J. (n.d.). *Screening tests*. Retrieved from Understanding Uncertainty: <https://understandinguncertainty.org/screening>.
- Sturm, A., & Eichler, A. (2014). Students' beliefs about the benefit of statistical knowledge when perceiving information through daily media. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9)*, Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute.
- Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, 1054–1059. <https://doi.org/10.1126/science.1196404>.
- Teigen, K. H., & Keren, G. (2007). Waiting for the bus: When base-rates refuse to be neglected. *Cognition*, 103, 337–357. <https://doi.org/10.1016/j.cognition.2006.03.007>.
- Thomas, M. O. (2008). Conceptual representations and versatile mathematical thinking. In *Proceedings of the Tenth International Congress in Mathematics Education, Copenhagen, Denmark* (pp. 1–18).
- Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from Bayes theorem and the additivity principle. *Memory & Cognition*, 30, 171–178. <https://doi.org/10.3758/BF03195278>.
- Ware, C. (2008). *Visual thinking for design*. Burlington, MA: Morgan Kaufmann Publishers.
- Watson, J. M., & 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. <https://doi.org/10.1080/10986065.2014.953019>.
- Wolfe, C. R. (1995). Information seeking on Bayesian conditional probability problems: A fuzzy-trace theory. *Journal of Behavioral Decision Making*, 8, 85–108.
- Zikmund-Fisher, B. J., Witteman, H. O., Dickson, M., Fuhrel-Forbis, A., Khan, V. C., Exe, N. L., et al. (2014). Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Medical Decision Making*, 34, 443–453. <https://doi.org/10.1177/0272989X13511706>.

Chapter 2

Students' Development of Measures



Christian Büscher

Abstract Knowledge is situated, and so are learning processes. Although contextual knowledge has always played an important role in statistics education research, there exists a need for a theoretical framework for describing students' development of statistical concepts. A conceptualization of measure is introduced that links concept development to the development of measures, which consists of the three mathematizing activities of structuring phenomena, formalizing communication, and creating evidence. In a qualitative study in the framework of topic-specific design research, learners' development of measures is reconstructed on a micro level. The analysis reveals impact of the context of a teaching-learning arrangement for students' situated concept development.

Keywords Concept development · Design research · Situativity of knowledge
Statistical measures · Statistical reasoning

2.1 Introduction: Concept Development as a Focus for Research

In recent years, the ability to draw Informal Statistical Inferences (ISI) (Makar and Rubin 2009) has become a focal point of statistics education research (see the ESM special issue on sampling, Ben-Zvi et al. 2015). ISI emphasizes the use of statistical concepts in drawing 'probabilistic generalizations from data' (Makar and Rubin 2009, p. 85) and in making claims about unknown phenomena. In order to describe the type of reasoning used in drawing ISI, Makar et al. (2011) propose a framework of Informal Inferential Reasoning (IIR). This framework reveals the complexity of IIR; the components include knowledge of statistical concepts as well as contextual knowledge and general norms, habits, and patterns of action (Makar et al. 2011).

C. Büscher (✉)
TU Dortmund University, Dortmund, Germany
e-mail: christian.buescher@math.tu-dortmund.de

© Springer Nature Switzerland AG 2019
G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_2

Learning to draw ISI is thus conceptualized by the development of IIR, placing strong emphasis on the *use* of statistical concepts in complex activity.

Learners however will first need to *develop* the statistical concepts to be used in IIR through activity in teaching-learning arrangements that are designed to facilitate such concept development. A key assumption of the research related to this study follows the ideas of Freudenthal (1991) that learners can develop formal, general concepts out of their informal, singular activity, when their learning processes are carefully guided. The framework of IIR gives only limited guidance for such a design, detailing the goal, but not the path, of concept development. A language is required that can be used to describe learners' situated understandings, the individual concepts guiding their actions, the relation of these concepts to formal statistical concepts, and the complex interplay with elements of the design of a teaching-learning arrangement.

Whereas similar studies have focused on the concepts of distribution (Lehrer and Schauble 2004) and shape (Gravemeijer 2007), this study focuses on the concept of measure. Since data-based evidence plays a major role in ISI and IIR and measures are a common form of such evidence, this focus could provide not only insights into concept development but also make connections to the development of IIR.

2.2 Theoretical Background

2.2.1 *The Situativity of Knowledge*

A long-standing perspective in cognitive psychology concerns the situativity of knowledge (Greeno 1998). A conceptualization of measure that takes learning processes into consideration needs to pay attention to the fact that knowledge emerges from situations. Vergnaud (1990, 1996) proposes a theory of conceptual fields as an epistemological framework. To Vergnaud, the perception of situations and the understanding of mathematical concepts stand in a dialectic relationship: "cognition is first of all conceptualization, and conceptualization is specific to the domain of phenomena" (Vergnaud 1996, p. 224). In this way, mathematical knowledge emerges through actions in situations. This knowledge is not to be understood as consisting of situation-independent abstractions but rather as an operational invariant across different situations.

The two most important types of operational invariants are concepts-in-action and theorems-in-action. Concepts-in-action are "categories (objects, properties, relationships, transformations, processes etc.) that enable the subject to cut the real world into distinct elements and aspects, and pick up the most adequate selection of information according to the situation and scheme involved" (Vergnaud 1996, p. 225). Thus, they organize what students focus on and in this case how they structure phenomena unknown to them. Theorems-in-action are defined as "propositions that [are] held to be true by the individual subject for a certain range of situation variables" (Vergnaud 1996, p. 225). They are intricately connected to the learners' concepts-in-action:

theorems-in-action give meaning to concepts-in-action, which in turn give content to the theorems-in-action.

A conceptualization of measure that takes into account the situativity of knowledge thus will need to provide a clear focus on the use of measures in situations. The median is a measure of center, but this does not explain its use in terms of operational invariants.

2.2.2 *Functions of Measures*

Although measures are a prominent concept in statistics and statistics education, few explicit definitions or conceptualizations exist that explain this construct. At least three different functions of measures can be identified based on literature: (a) structuring phenomena, (b) formalizing communication, and (c) creating evidence.

Structuring phenomena. Bakker and Gravemeijer (2004) distinguish between data (the individual values) and distribution (a conceptual entity). Two perspectives on data and distribution emerge: The 'upward perspective' consists in seeing data as a means to calculate measures (median, range, ...) of aspects of a distribution (center, spread, ...). The 'downward perspective' consists of looking at the data from the standpoint of distribution, with aspects of center and spread as organizing structures already in mind.

In this way, measures function as lenses that allow access to distributional properties. This resonates with the idea of an 'aggregate view' on data (Konold et al. 2015): perceiving data as a conceptual unit with its own emergent properties, which can be accessed through the use of measures. In data investigation, measures thus impose distributional properties on phenomena, creating structure in previously unstructured phenomena.

Formalizing communication. Structuring phenomena alone does not conclude statistical investigation; findings must also be communicated to a wider audience. Through their standardized procedures of calculation, measures can provide such a means of communication. They create intersubjectivity, allowing for communication about phenomena across distance and time (Porter 1995; Fischer 1988).

Creating evidence. One of the characterizing features of ISI given by Makar and Ruben (2009) is the use of data as evidence. Whereas they do not explicitly relate this role of evidence to measures, it is possible to think of the form of this evidence as consisting of measures. Abelson (1995) states that the discipline of statistics supports principled arguments that aim at changing beliefs and which therefore need to be convincing to others. Simple unspecified reference to data would not serve this goal of convincingness. Instead, specific aspects have to be 'singled out' that explicate what exactly is convincing in the data. This is a role played by measures.

2.2.3 A Conceptualization of Measure

Although the list of functions of measures presented above possibly is not complete, it illustrates some common facets of the use of measures on which each function places different emphasis, in different terms, with varying grades of explicitness.

Measures are *grounded in data*. Although this facet on its own is not terribly surprising, the role of measures becomes clearer when related to another facet: measures *describe phenomena*. They bridge the gap between data and phenomenon. A phenomenon behind some data can be accessed through the use of measures that operate on that data. This can lead to new insights into the phenomenon and is a prerequisite for communication about that phenomenon. Measures however can never capture the full phenomenon but provide *discrete* descriptions. They separate phenomena into relevant and irrelevant parts, highlighting only very specific *aspects* of phenomena. This is the reason why they can provide convincing principled arguments and give new, but also possibly incomplete, insights into phenomena.

From these considerations, this study draws a conceptualization for the concept of measure: a measure is a *data-based description of one aspect of a phenomenon*. This definition builds on a broad understanding of the term ‘phenomenon’. ‘Aspect of a phenomenon’ can refer to any part of a phenomenon that is held to be relevant for a specific question in a specific situation. An example could be the daily ice growth used by climate scientists as a measure of the volatility of the melting process of Arctic sea ice (Fig. 2.1). Another aspect of the same phenomenon could be the general well-being of the Arctic ice sheet, addressed through the measure of monthly average extent (Fetterer et al. 2002). While these aspects are phenomenon-specific, measures can also refer to more general aspects like the central tendency. A distinction can be drawn between *general measures* that focus on *general* aspects of phenomena

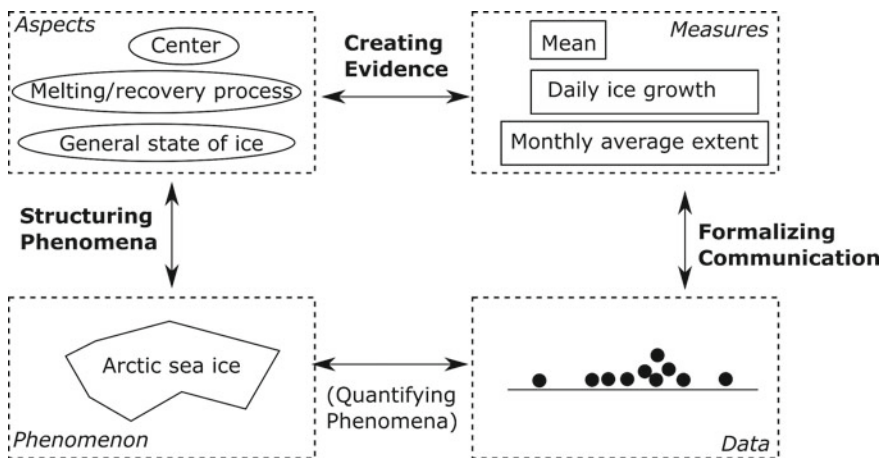


Fig. 2.1 The relations between phenomenon, data, measures, and aspects of phenomena

(center, spread, ...) and *situative measures* that focus on phenomenon-specific *situative* aspects. General measures consist of all measures commonly referred to in formal statistics, whereas situative measures address phenomenon-specific aspects such as the melting process of Arctic sea ice. The meaning of situative measures is often situation-specific, whereas general measures provide situation-independent tools for structuring phenomena. This does not mean that the use of general measures is strictly situation-independent: general measures can also be used to address phenomenon-specific aspects.

2.2.4 *The Development of Measures*

Whereas statisticians are able to use general measures such as the median to address the center of arbitrary phenomena, the situated nature of knowledge implies that learners will have to resort to phenomenon-specific situative measures when starting out in their learning trajectory. This puts the learners into an inconvenient position. They will need to structure phenomena by identifying aspects, while simultaneously finding situative measures to address just these aspects. Learners need to *develop* their measures. During their learning process, learners will need to answer questions corresponding to a measures' functions of structuring phenomena, formalizing communication, and creating evidence.

As emphasized by Vergnaud (1996), a theory of learning needs to give a prominent place to learners' activities. In order to illustrate how formal ideas can emerge from informal activity, the functions of measures are now (in reference to Freudenthal 1991) interpreted as *mathematizing activities* carried out by the learners while developing measures. When engaging in the mathematizing activity of structuring phenomena, learners make sense of a situation through their concepts-in-action. Their contextual knowledge of the phenomenon plays an important part, as they have not yet developed general measures for structuring phenomena. The mathematizing activity of formalizing communication focuses on a measure's formal characteristics, such as definition, calculation, and rules of application. In the beginning of learning processes, visual identification (i.e. 'just seeing') would be an adequate way of finding an situative measure. However, such visual identification could hardly provide intersubjectivity; finding standard procedures of calculation instead could be an act of formalizing communication. During the mathematizing activity of creating evidence, learners decide the actual aspects and measures to be chosen for argumentation. Again, contextual knowledge can play an important part for clarifying which aspects are relevant for which questions regarding the phenomenon and thus, which line of argumentation should be supported by what evidence.

Through the investigation of different phenomena, operational invariants over different situations can emerge, making the use of situative measures less phenomenon-dependent. In this framework, learning takes the form of developing situative measures into general measures through mathematizing activity across

different situations: broadening the aspects of phenomena addressed by measures, explicating formal characteristics, and supporting argumentation through evidence.

2.2.5 *Research Questions*

The starting point for this study was the need for a conceptualization of measure that allows for the design of a teaching-learning arrangement that draws on learners' situated understandings and can lead to the development of statistical concepts. Such a teaching-learning arrangement needs to elicit the mathematizing activities of structuring phenomena, formalizing communication, and creating evidence. Since the role of those activities was based on theoretical observations, it remains unclear how actual learning processes are constituted in these activities and how a teaching-learning arrangement can support them. Although all three mathematizing activities play a part in the development of measures, this study limits itself by focusing on the activities of structuring phenomena and formalizing communication in order to provide a more in-depth view of the learning processes. The empirical part of this study thus follows the following research questions:

- (RQ1) How can design elements of a teaching-learning arrangement elicit and support the mathematizing activities?
- (RQ2) How do learners' situative measures develop through the mathematizing activities of structuring phenomena and formalizing communication?

2.3 **Research Design**

2.3.1 *Topic-Specific Didactical Design Research as Framework*

Design research as methodological frame

The presented study is part of a larger project in the framework of topic-specific didactical design research (Prediger et al. 2012). This framework simultaneously aims at two different but strongly interconnected goals: empirically grounded theories on the nature of topic-specific learning processes and learning goals (i.e. what and how to learn), and design principles and concrete teaching-learning arrangements for learning this topic (i.e. with what to learn). This is achieved by a focus on learning processes (Prediger et al. 2015). Special attention is given to the careful specification and structuring of the learning content as well as to developing content-specific local theories of teaching and learning (Hußmann and Prediger 2016).

Research is structured into iterative cycles consisting of four different working areas (see Fig. 2.2). In a first working area, the learning content is specified and struc-

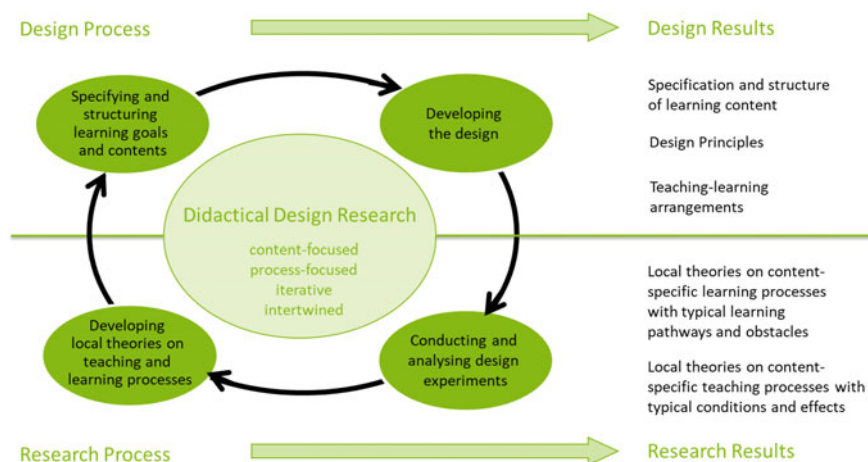


Fig. 2.2 The cycle of topic-specific didactical design research (Prediger et al. 2012; translated in Prediger and Zwetzschler 2013, p. 411)

tured, identifying central insights into the content that learners need to achieve and structuring them into possible learning pathways. This can be based on epistemological considerations such as a didactical phenomenology (Freudenthal 1983) as well as on empirical insights into possible learning obstacles and students' conceptions. The second working area consists of designing a teaching-learning arrangement to be used in the third working area, conducting design experiments (Cobb et al. 2003). The learning processes initiated in the design experiments are then analyzed and serve as a basis in developing local theories about these teaching and learning processes. A main strength of the framework of didactical design research is the interconnectedness of these working areas: in the next cycle, the local theories developed can inform the re-specification and re-structuring of the learning content. This re-structuring in turn influences the design principles enacted in the teaching-learning arrangement and thus, the initiated learning processes. Through this process, theory and design get successively more refined in each cycle.

2.3.1.1 Participants and Data Collection

This study reports on findings of the third cycle of design experiments of the ongoing design research project (for other results see Büscher 2017, 2018; Büscher and Schnell 2017; Schnell and Büscher 2015). The design experiment series in the third cycle took place in laboratory settings with five pairs of students in a German middle school (ages 12–14). Each pair took part in a series of two consecutive design experiment sessions of 45 min each. The participating students were chosen by their teacher as performing well or average in mathematics, which includes statistics education in German curricula. At the time of the experiments, the students had very little expe-

rience with statistics besides learning simple measures such as the arithmetic mean and median a year before in grade 6. They were familiar with frequency distributions but only on a rather superficial level (e.g. reading out information on maximum and minimum), without comparing them strategically. They were not familiar with stacked dot plots or measures of spread.

All experiments were completely videotaped (altogether 450 min of video in the third cycle). Here, the case of two pairs of students is presented, selected due to the richness of their communication and mathematizing activities. Their design experiment sessions were fully transcribed.

2.3.1.2 Data Analysis

The qualitative data analysis aims not at solely assigning students' utterances to the general statistical concepts but instead at capturing the individual emergent, situative concepts. In order to capture the richness and heterogeneity of the students' individual reasoning, this study chose a category-developing approach (cf. Mayring 2000) using open and interpretative approaches (cf. Corbin and Strauss 1990) for identifying individual concepts-in-action and theorems-in-action (Vergnaud 1996) based on the students' utterances and gestures. This methodological foundation of the analytical framework by Vergnaud's constructs allow the data analysts to capture the situativity of knowledge and learning. The identified individual concepts-in-action and theorems-in-action on measures are not necessarily in line with general statistics concepts but rather mirror their own situative structure of phenomena. In the analysis, concepts-in-action are symbolized by $\| \dots \|$ and theorems-in-action by $\langle \dots \rangle$.

2.3.2 Design Principles

During the five design experiment cycles, several design principles were implemented and iteratively refined that played a role in initiating concept development. Three design principles play an important part in this study (for a complete overview see Büscher 2018); each of the design principles focused on eliciting a different mathematizing activity.

Investigating realistic phenomena. A teaching-learning arrangement focusing on the development of measures needs to elicit the mathematizing activity of structuring phenomena. Since most students do not yet have access to phenomenon-independent measures to structure arbitrary unknown phenomena, the choice of the phenomenon to be investigated has to be carefully considered. This study uses phenomena such as variability in the weather that are close enough to students' reality so that they can informally and intuitively structure the phenomenon.

Scaffolding the use of measures in argumentation. Previous cycles of the project showed how students did use situative and occasionally even general measures when comparing distributions. Whereas there was a lot of potential in this, their uses stayed elusive: students lacked the language to specify the addressed aspects and formal characteristics of their situative measures—they struggled to formalize their communication. This led to an insecure use of measures, so that they sometimes simply had already forgotten their train of thoughts when prompted by the researcher or other students. This raised the need to scaffold the use of measures by explicating their use in giving arguments about phenomena. This design principle was implemented through the use of so-called report sheets (see below)

Contrasting measures. Central to the measure-focused approach of this study is the insight that different measures for the same distribution can result in different views on the situation by emphasizing different aspects. Thus, engaging in the activity of creating evidence can mean to contrast and evaluate different measures with respect to (a) their usefulness regarding specific investigations, (b) their correspondence to learners' experienced reality, (c) their applicability in different situations, or (d) their advantages or disadvantages in argumentation. This design principle was realized by contrasting different report-sheets (see below).

2.3.3 Task Design

The design of the two sessions of the design experiments consisted of two different tasks, the *Antarctic weather task* (Session I) and the *Arctic sea ice task* (Session II). Each task was structured into different phases, with progressions between phases initiated by the researcher when certain requirements were met.

The Antarctic weather task

The goal of this task was to introduce the students to the idea of measures and the design elements central to the whole design experiment. The task was structured into three phases.

Phase I.1. The students were given dot plots of temperature distribution at the Norwegian Antarctic research station *Troll forskningsstasjon* (Fig. 2.3, data slightly modified from Stroeve and Shuman 2004) and introduced to the setting of the task: as consultants to researchers planning a trip next year, they were charged with giving a report of the temperature conditions. Since the students were unfamiliar with dot plots, special attention was given to make sure that students understood the diagrams. The data were presented to the students on a tablet with a screen overlay software to allow for drawing visualizations of their situative measures directly onto the screen. *Tinkerplots2.0* (Konold and Miller 2011) was used to create the diagrams, without giving the students access to interactive functionalities of the software. When sufficient understanding of the diagrams was achieved and the students had given some informal predictions for next year, the task progressed to Phase I.2.

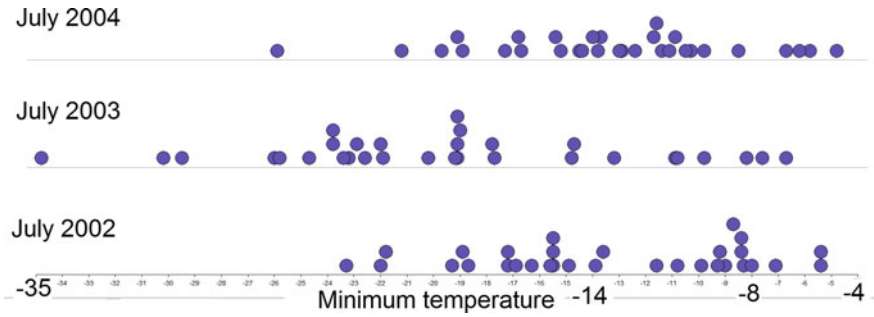


Fig. 2.3 Distributions of the Antarctic weather task (translated from German)

Fig. 2.4 Empty report sheet (translated from German)

| Report sheet: Temperatures at Troll Forskningsstasjon | |
|---|----------------|
| Sketch | Summary |
| | |
| Minimum: | |
| Maximum: | |
| Typical: | |

Phase I.2. In this phase, the students were introduced to the central design element of the design experiments, the *report sheets* (Fig. 2.4). These report sheets served as a scaffold for argumentation with measures, combining a graphical representation with measures and a brief inference about the phenomenon of Antarctic weather. The students were asked to fill out a report sheet to be used as a report for the researchers. The measures employed were given to them without explanation, so that they needed to find their individual interpretation of minimum, maximum, and typical. Typical here served as an situative measure for a yet unspecified situative aspect, which could be interpreted by the students as incorporating some aspect of variability (similar to Konold et al. 2002) Since formal characteristics and the meanings of the measures were left unspecified, this task aimed at eliciting the mathematizing activities of structuring phenomena and formalizing communication.

Phase I.3. After the students had created their own report sheet, they were given fictitious students' *filled-in report sheets* (Fig. 2.5). These report sheet differed in their interpretations of the measures employed and thus focused on different aspects of the phenomenon. The students were asked to evaluate these report sheets and to possibly adapt their own report sheet.

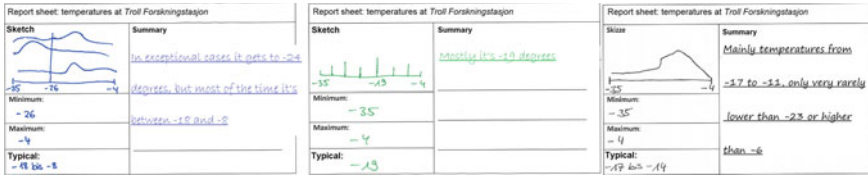


Fig. 2.5 Fictitious students' filled-in report sheets (translated from German)

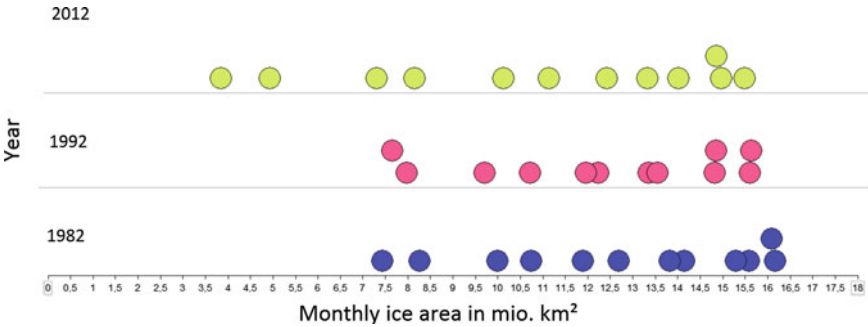


Fig. 2.6 Distribution of sea ice extent in Session II (translated from German)

The Arctic sea ice task

Phase II.1. The Arctic sea ice task followed a similar progression as the Antarctic weather task. This time the students were put into the roles of researchers of climate changes. Students were given distributions of monthly lowest Arctic sea ice extent for the years 1982, 1992, and 2012 (Fig. 2.6; data slightly modified from Fetterer et al. 2002) and were asked to give a report whether, and how much, the ice area had changed. This phase again aimed at ensuring the students' understanding of the diagram and the context. They were not yet asked to create a report sheet.

Phase II.2. Following the introduction of the setting, the students again received filled-in report sheets (Fig. 2.7). These report sheets now allowed for arbitrary measures and again presented different formalizations of measures and abstractions of the phenomenon. This time the different measures lead to radically different perceptions of the phenomenon of Arctic sea ice, with report sheets proclaiming either no change or radical change in Arctic sea ice (Fig. 2.7). Discussion revolved around which report sheet was right, and what a researcher would need focus on when reporting on Arctic sea ice, thus eliciting the mathematizing activity of creating evidence.

Phase II.3. Following the discussion, the students were again asked to create their own report sheet. Whereas the students were free to choose their measures for the report sheet, the students were expected to adapt elements of the filled-in report sheets for their own report sheet. This initiated further mathematizing activities, as the students were asked to justify their choice of measures.

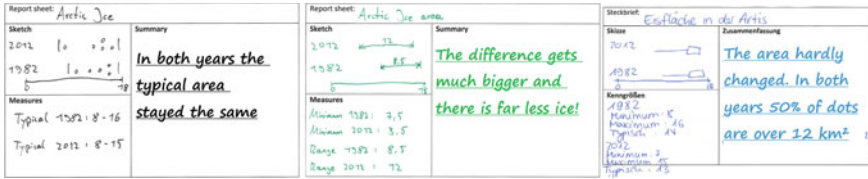


Fig. 2.7 Filled-in report sheets for Session II (translated from German)

2.4 Empirical Results

This study identifies students’ mathematizing activities to investigate their developing measures. The first part of this section follows the learning processes of two students, Maria and Natalie, through both sessions of the design experiment. Due to the rapid changes of the roles in the students’ interaction, the transcript has been partially cleaned up to increase readability. The analysis focuses on their use of the situative measure of Typical (reference to the situative measure Typical indicated by capital-T), from its unspecified beginning in Session I to its more formalized version at the end of Session II. During the design experiment, the students get increasingly precise in addressing different aspects of phenomena and in structuring the phenomenon. This is then briefly contrasted with the processes of another pair, Quanna and Rebecca, focusing on Session II and highlighting similarities and differences in the two pairs’ use of Typical.

2.4.1 The Case of Maria and Natalie

Session I: The Antarctic weather task

The first snapshot starts with Phase I.1 of the Antarctic weather task. After giving some informal predictions of the weather, Maria (M) and Natalie (N) try to explicate their view on the data to the researcher (I).

- 1 M We are pondering what the relationship, like, how to...
- 2 N Yes, because we want to know what changes in each year. And we said that there [points to 2003] it came apart.

[...]

- 8 M Yes, I think it [points to 2004] is somehow similar to that [points to 2002], but that one [points to 2003] is different.
- 9 N Like here [points to 2004, around -12 °C] are, like, like the most dots, and here [points to 2002, -12 °C] are almost none. And there [points to 2002, -8 °C] are the most and here [points to 2004, -8 °C] are almost none.

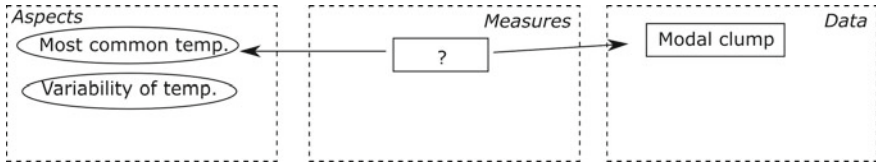


Fig. 2.8 Maria and Natalie’s use of measures (Part 1)

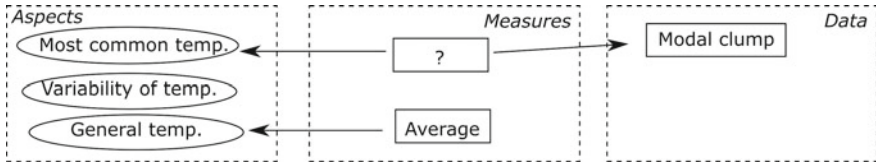


Fig. 2.9 Maria and Natalie’s use of measures (Part 2)

This excerpt serves as an illustration of the starting point in the students’ reasoning. The students are trying to characterize the differences observed in the distributions. In order to do this, they structure the phenomenon by identifying two aspects: the *most common temperatures* (where “the most” temperatures lie, #9), and the *variability of temperatures* (how they “came apart”, #2). Whereas the students are able to use modal clumps as a way to address the *most common temperatures*, they seem to lack ways of addressing the *variability of temperatures* (Fig. 2.8).

A few minutes later, the students find a way to better address the difference between the distributions.

21 M Well, we first should look at how many degrees it has risen or fallen. Generally. In two years.

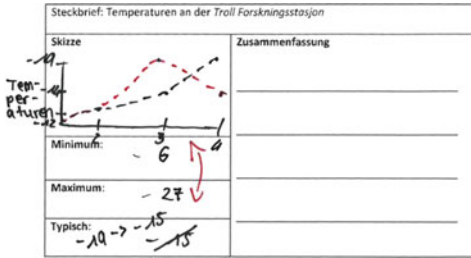
[...]

27 N You mean average, like...

28 M The average, and then we look at how the average changed in two years.

By identifying the aspect of a *general temperature* (“Generally”, #21), the students are able to re-structure the phenomenon to reduce the complexity of the temperatures. For this aspect, they appear to already know an adequate general measure: the *average*. To the students, *<the average addresses the general temperature of a distribution>*. This *general temperature* does not necessarily correspond to the *most common temperatures* addressed earlier. In this way, the phenomenon gains additional structure (Fig. 2.9).

The design experiment progresses through Phase I.2, in which the students create their own report sheet (Fig. 2.10). The analysis picks up at beginning of Phase I.3, with the students comparing the different interpretations of Typical in the filled-in report sheets.



Translations:
 "Report sheet: temperatures at Troll Forschungsstation"
 Skizze – Sketch
 Typisch – Typical
 Zusammenfassung – Summary
 Temperaturen - Temperatures
 (The black graph was drawn first, labeled a mistake, and immediately replaced by the red graph. Typical was first assigned as -19, then during Phase I.3 changed to '-19 to -15'.)

Fig. 2.10 Maria and Natalie’s first report sheet

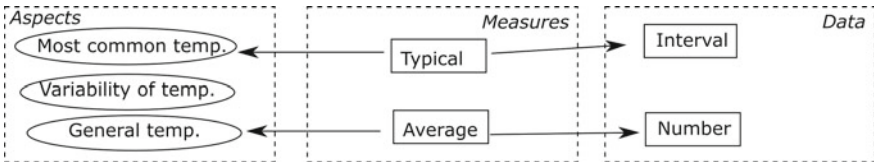


Fig. 2.11 Maria and Natalie’s use of measures (Part 3)

Comparing the different interpretations of Typical, Maria and Natalie are intrigued by the possibility to use an interval to formalize Typical. This consideration leads them to reflect on their use of the average.

- 41 N But the average temperature isn’t really typical, is it?
- 42 M What, typical? Of course the average temperature is the typical.
- [...]
- 46 M Well, no. Typical is more like where the most... no...
- 47 M The average temperature isn’t the typical after all. Because it’s only the general, the whole. The typical would be for example for this [2004] here [points to -14 on the 2004 dot plot].
- 48 N Typical I think simply is what is the most or the most common.

The students differentiate between average and Typical to address different aspects: The *general temperature* is addressed by the general measure *average* (“the general, the whole”, #47), and the *most common temperatures* addressed by situative measure *Typical* (“the most common”, #48). At this point it is not yet clear if the situative measure Typical consists of a number or an interval—it is still in need of formalization. However, introduction of this situative measure seems to have allowed the students to reconnect to the aspect of *most common temperatures* (first expressed in #9) that got swept aside by the more formalized average (Fig. 2.11).

Some minutes later, Natalie summarizes her view on the relation between Typical and average.

- 61 N Average is pretty imprecise, because it doesn't say anything about a single day. And with Typical, I'd say, that it's a span between two numbers, because that way you can better overlook how it is most of the time.

In the end of Session I, the measures Typical and average address two different aspects of the phenomenon of Antarctic weather. Whereas the average addresses the general temperature, Typical describes the most common temperatures. The average can be used to compare distributions, whereas Typical gives an insight into a range of 'normal' or 'expected' temperatures, to which any single day can be compared. Central to this distinction was the formalization of Typical as an interval.

Session II: The Arctic sea ice task

Most of Session II revolved around the question how to further formalize Typical, and how to distinguish it from the average. This excerpt starts in the middle of Phase II.2, and takes place over a period of eight minutes. In the preceding minutes, the students had used the average to propose a general decline in Arctic sea ice.

- 1 I Last time we talked about Typical, and here Typical is also drawn in. Do you think that's helpful, or not?
 2 M Typical, wait a second, there [report sheet 3, 1982] Typical is 14 right? Huh, but why is 13 Typical here [report sheet 3, 2012]?
 3 N Huh, Typical can't be 13, because Typical actually is a range, isn't it?

[...]

- 6 I What would you say what one should choose?
 7 N I would definitely say a range, because that just tells you more. Because you can't say that it's 11 degrees typical.

Maria and Natalie are irritated by the same report sheets showing different values for the measure *Typical* (#2). This leads them to question whether Typical should be formalized as a number or an interval ("range", #3).

In Session I, the students opted for the interval. Natalie draws on this knowledge, postulating that *Typical cannot be a number, because numbers do not describe Typical temperatures* ("you can't say that it's 11 degrees typical", #7). In this way, she uses the situative aspect of *most common temperatures* from Session I to formalize the situative measure Typical in another phenomenon as an interval. Because this transfer of phenomena happens frequently throughout the session (see below), it could be seen as the emergence of operational invariants across situations, rather than a simple mistake in wording.

Some moments later, after they have again considered the average, the students compare the two measures.

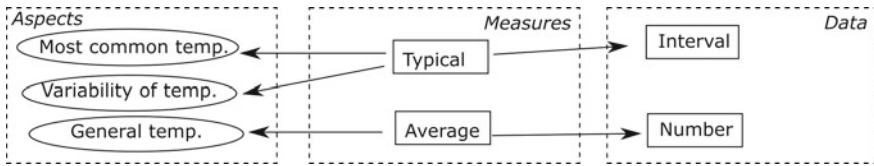


Fig. 2.12 Maria and Natalie’s use of measures (Part 4)

- 21 I And if you would create such a report sheet, would the average suffice?
 22 N No. Well I think the average is important, isn’t it? But a range, what’s typical, that just tells you more about single days than if you take the average.
 23 N Because if the average is like 12, then one day could be 18 degrees, or –10 degrees or something. And the average better tells you what generally happened, and I think a range better tells you what happened generally.
 24 N Because if the average was 8 degrees, but it also happened to get to 18 degrees or –10 degrees, then the range would rather be from 5 degrees to – I don’t know.

Natalie distinguishes between two aspects: what “generally happened”, and what “happened generally” (#23). These are two different (yet unnamed) aspects, because the distinction serves as an explanation of the distinction between average and Typical (sometimes referred to by Natalie as “range”, #23). Natalie seems to lack the vocabulary to clearly differentiate between the two aspects. In her explanation however she again seems to draw on an aspect of the previous session: the *variability of temperatures*, as she states that *<a high variability of temperatures can be seen in the Typical range>* (temperatures from 18 to –10 would somehow be reflected in the “range”, #24), whereas *<the average is not impacted by the variability of temperatures>* (the average would stay at 8 degrees, #24). Again, the formalization of the measure Typical progresses by drawing on the structuring of another phenomenon (Fig. 2.12).

Following this exchange, after some minutes, the students return to the problem of finding the Typical interval.

- 41 N I don’t know how to calculate Typical. I think you start from the average, and then looks at the lowest and highest temperatures, and from that you take a middle value. Like between the average, and...
 42 M And the lowest and the highest... we are talking about temperatures the whole time, but those aren’t temperatures.
 43 N Yes but if we took temperatures, then you take the average and the coldest and then again take the average.

[...]

- 48 N And then the average from the average is the Typical. Between this average and that.

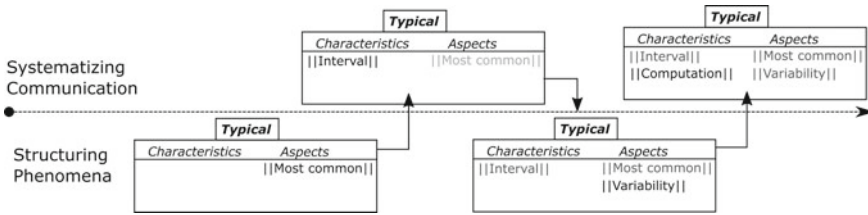


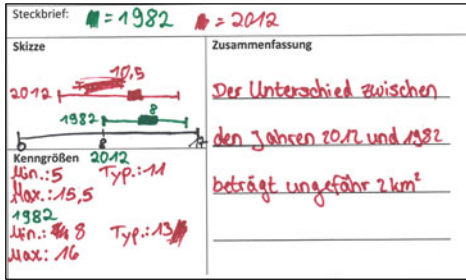
Fig. 2.13 Maria and Natalie’s development of Typical

With the aspects addressed by the two measures now firmly separated, the students find a way to calculate their Typical interval: Taking the average of the whole distribution (“start from the average, #41), splitting the distribution into two halves at this point (“look at the lowest and highest temperatures, #41), calculating the average for each of those halves (“again take the average, #43), and then taking the interval between those two averages (“between this average and that”, #48). This shows a highly formalized use of the average: the *||general temperature||* addressed by the measure *||laveragell* seems to also apply to only halves of distributions. This excerpt is the first time one of the students becomes aware of their substitution of the phenomenon of Arctic sea ice with Antarctic temperatures (#42). The casualness of Natalie’s dismissal of this fact (“yes but if we took temperatures”, #43) however seems to suggest that the operational invariants of Typical in the end seems to encompass both situations, temperatures and sea ice.

Summary. During Session I, Maria and Natalie structure the phenomenon of Antarctic temperatures into *||most common temperatures||*, *||general temperature||*, and *||variability of temperatures||*. They also determine formal characteristics of the measure Typical by formalizing it as an interval, in contrast to the average. This distinction is transferred to another phenomenon in Session II, but not without problems: again, the characteristic of Typical as an interval must be justified. In the end, the students even arrive at a way of finding the Typical interval that is similar to that of finding the interquartile range. During the whole learning process, the situative measure of Typical develops in interrelated mathematizing activities of structuring phenomena and formalizing communication. Figure 2.13 provides an overview on this development.

2.4.2 The Case of Quanna and Rebecca

The following empirical snapshot follows the students Quanna (Q) and Rebecca (R) in Session II of the design experiment. The excerpts stem from a conversation of about 15 min. The snapshot starts in Phase II.3 with the students filling out their own report sheet (shown in Fig. 2.14, but not completed until turn #40) after they have discussed the filled-in report sheets.



Translations:
 Steckbrief: - Report sheet:
 Skizze – Sketch
 Typ. – Typical
 Zusammenfassung – Summary: “The difference between the years 2012 and 1982 is about 2 km²”
 Temperaturen - Temperatures
 (Typical in 1982 was later changed from 13.5 to 13)

Fig. 2.14 Quanna and Rebecca’s report sheet

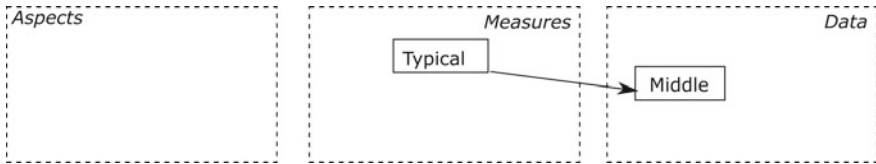


Fig. 2.15 Quanna and Rebecca’s use of measures (Part 1)

- 1 Q [while filling out own report sheet] And Typical...
- 2 R Typical [...] it could be, like, the middle or something?
- 3 R I would say the middle and a bit higher.

Although they could have referred to other measures, *Typical* is the main measure organizing their view on the phenomenon. Without paying attention on the aspects to be addressed, the students are formalizing *Typical* as located in the *middle* of the distribution: *<Typical is located a bit higher than the middle>* (#3) (Fig. 2.15).

Some minutes later, the students are about to write their summary for the report sheet.

- 20 Q Okay, now the summary.
- 21 R The numbers got [points to own report sheet] – look – more ice melted away.
- 22 Q [shakes head] the difference is – is around 2.5.
- 23 R Always?
- 24 Q Yes, right here [points to own report sheet] of Typical.

Rebecca seems to have difficulties with combining the phenomenon (the melting ice) with the task of giving a short data-backed summary. At this point, Quanna is able to utilize their measure of Typical. In the meantime, the students had decided that *<Typical is a number>*, which they intuitively identified for the distributions of 1982 and 2012 as 11 and 13.5. These numbers show a difference of 2.5, which can now be used in their summary to report on the Arctic sea ice decline: *<Typical*

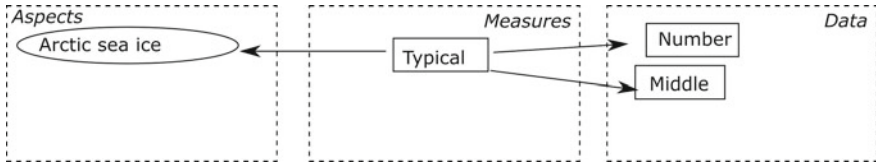


Fig. 2.16 Quanna and Rebecca’s use of measures (Part 2)

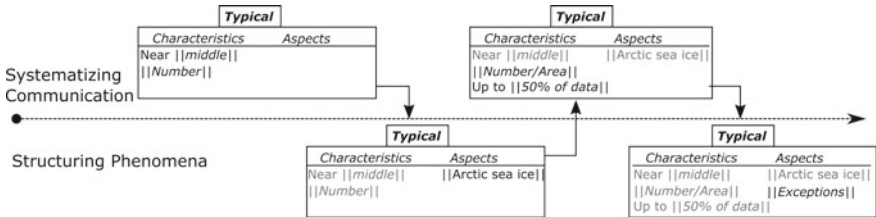


Fig. 2.17 Quanna and Rebecca’s development of Typical

can be used to address the state of Arctic sea ice> (the melting ice, #21, addressed through the difference of Typical, #22). However, it remains unspecified what exactly is meant by this aspect of a general *state of Arctic sea ice* (Fig. 2.16).

The characteristics of Typical still being unclear, the researcher challenges them to explain their use of Typical.

- 40 I I see you decided to use only one number for Typical, in contrast to this report sheet, where they used an area [points to filled-in report sheet]. Is that better or worse, what do you think?
- 41 R Well Typical is more of a single...
- 42 Q [simultaneously] more of an area...
- 43 R Now we disagree. [...] Typical is more of a small area, or you could say a number. Like here, from 10 to 12. [...] If the area is over 100, it may be over 10. [...] But never more than the half.

The claim *Typical is a number* becomes disputed, as *number* and *area* both are possible characteristics of Typical, as evidenced by the filled-in report sheet in Fig. 2.17. This initiates further processes of formalization, resulting in more pronounced formal characteristics of Typical. Whereas there still is no full definition, there are criteria for its correct form: *Typical is an area that at most covers half the data* (“never more than the half”, #43) and *Typical can be signified by a number, if the area is small* (“small area, or you could say a number”, #43).

Some minutes later in the discussion, Rebecca tackles the question whether one is allowed to omit data points that could be seen as exceptions when creating report sheets.

- 61 R Well, you can do that, but it depends. You have to make sure it fits. If you do it like here [*points to own report sheet*] you should not consider the isolated cases [...] because then it gets imprecise. But if the Typical area was the same on both sides, I think you can do that.

Whereas there still is no full definition of Typical, another situative aspect has been added that is addressed by Typical. Typical not only functions as a description of a general *||state of Arctic sea ice||*, but also addresses *||rule and exceptions||* of the Arctic ice: *<If the Typical area of two distributions is the same, one can use Typical to address exceptions>*. In this way, the formalization of Typical as an interval in the middle of the distribution has allowed for addressing a previously unstructured aspect of the phenomenon.

Summary. Throughout this episode, the students expand the aspects addressed through Typical as well as the situative measures' formal characteristics. In the end, they use Typical to address a wide range of aspects that could also be addressed through general statistical measures (Fig. 2.17). The differentiation of aspects of phenomena and the growing explicitness in formal characteristics of Typical took place in interlocking mathematizing activities of structuring phenomena and formalizing communication: after Typical has become sufficiently formalized, it could be used to structure the phenomenon of Arctic sea ice into *||rule and exceptions||*.

2.5 Conclusion

This study started out from the need for a conceptualization of learners' situated understandings and the development of statistical concepts through their activities in learning processes. The concept of measure was introduced, distinguishing between general and situative measures: measures that address phenomenon-specific aspects without necessarily showing explicit formal characteristics. Learning took place during the development of learners' situative measures through the three mathematizing activities of structuring phenomena, formalizing communication, and creating evidence. An empirical study was then used to illustrate (a) how learning processes can be understood through this conceptualization of measure and (b) how the design of a teaching-learning arrangement can influence these learning processes.

2.5.1 The Development of Measures

The analysis shows the students to be fully engaged in the mathematizing activities, which presented themselves as being intricately connected. Structuring phenomena into aspects provided the reason for formalizing the measures, and additional formal characteristics found for measures initiated further structuring of the phenomenon.

The more the students formalized their measure, the more situations were included in the operational invariants of the measure.

The interpretative approach to the analysis revealed the phenomenon-specificity of the students' measures. Maria and Natalie did not use the general measure of the average to address a general aspect of center, but a situative aspect of general temperature. This was then contrasted with the situative measure Typical, which was used to address a range of expected temperatures. Using Typical to structure the new phenomenon of Arctic sea ice took explicit reference back to the phenomenon of Antarctic temperatures. In this way, the students' knowledge of the structure of phenomena influenced their development of measures. They did not simply use measures to make sense of phenomena, but knowledge of situation and measure emerged at the same time.

One strategy that emerged for Maria and Natalie was the comparison of measures with differing degrees of formalization. Because the students knew the formal characteristics and aspects addressed by the general measure average, they could use it to develop the situative measure Typical. The average could even be employed in the calculation of Typical, leading to a measure that addressed aspects that could not adequately be addressed previously.

One idea postulated in the framework was the possibility of development of situative into general measures. Although the learning processes investigated in this study ended before the development of general measures, the findings suggest that this would indeed be possible. Both pairs of students ended with an situative measure Typical that resembled the general measure of the interquartile range. Quanna and Rebecca used Typical to describe an area in the middle of the distribution, consisting of no more than half the data points, indicating the location of the densest area, partitioning the distribution into rule and exception. Maria and Natalie calculated Typical by finding multiple averages, which would have resulted in the interquartile range had the average been substituted by the median. In their combination of average and Typical, Maria and Natalie manage to coordinate different measures, showing the possibility of creating understanding even for conceptually rich representations such as boxplots (cf. Bakker et al. 2004).

2.5.2 Supporting Mathematizing Activities

Central to the design of the teaching-learning arrangement was the choice of phenomenon to be investigated. Theoretical considerations led to the design principle of choosing realistic phenomena to be investigated. The choice of Antarctic weather and Arctic sea ice proved to be a fruitful one: in the case of Maria and Natalie, the students could identify aspects of phenomena regarding the natural variability and central tendency of weather. Through identification of operational invariants across the phenomena, the corresponding measures could then be broadened to also structure the phenomenon of Arctic sea ice.

Another design principle was the scaffolding of the use of measures in argumentation through the report sheets. This provided the students with different situative measures both pairs could appropriate for their individual reasoning. Since these measures were provided without explanation, and with different formal characteristics, the students needed to choose and commit to certain characteristics. In this way, this design principle of contrasting models led to the activity of formalizing communication.

2.5.3 *Limitations and Outlook*

With the study limited by its own situativity in the design of the teaching-learning arrangement and number of students analyzed, careful consideration has to be given to the generalizability of the results. The investigated development of measures has to be understood in the context of the design: the mathematizing activities were influenced by design elements, students, and the researcher. Any change in these factors could result in very different learning processes.

Yet the nature of this study was that of an existence proof of concept development and an illustration of a theoretical concept revealing a richness within the students' learning processes. Aiming for ecological validity (Prediger et al. 2015), this richness observed with only two pairs of students calls for analysis of additional pairs. Some results already indicate a wealth of strategies and conceptions, along with similarities in the development of measures (Büscher 2017, 2018; Büscher and Schnell 2017).

The analysis also showed the importance of the phenomenon not only as a motivating factor, but as integral to concept development itself. Further research could also be broadened to include other phenomena to be investigated. Since the learning processes was bound to the phenomena, a task design that focuses on other phenomena than weather and ice could provide other starting points for the development of measures.

References

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Erlbaum.
- Bakker, A., Biehler, R., & Konold, C. (2004). Should young students learn about boxplots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education* (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: Springer Netherlands.
- Ben-Zvi, D., Bakker, A., & Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics*, 88(3), 291–303.
- Büscher, C. (2017, February). *Common patterns of thought and statistics: Accessing variability through the typical*. Paper presented at the Tenth Congress of the European Society for Research in Mathematics Education, Dublin, Ireland.

- Büschler, C. (2018). *Mathematical literacy on statistical measures: A design research study*. Wiesbaden: Springer.
- Büschler, C., & Schnell, S. (2017). Students' emergent modelling of statistical measures—A case study. *Statistics Education Research Journal*, 16(2), 144–162.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.
- Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1), 3–21.
- Fetterer, F., Knowles, K., Meier, W., & Savoie, M. (2002, updated daily). *Sea ice index, version 1: Arctic Sea ice extent*. NSIDC: National Snow and Ice Data Center.
- Fischer, R. (1988). Didactics, mathematics, and communication. *For the Learning of Mathematics*, 8(2), 20–30.
- 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.
- Gravemeijer, K. (2007). Emergent modeling and iterative processes of design and improvement in mathematics education. In *Plenary lecture at the APEC-TSUKUBA International Conference III, Innovation of Classroom Teaching and Learning through Lesson Study—Focusing on Mathematical Communication*. Tokyo and Kanazawa, Japan.
- Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53(1), 5–26.
- Hußmann, S., & Prediger, S. (2016). Specifying and structuring mathematical topics. *Journal für Mathematik-Didaktik*, 37(S1), 33–67.
- 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., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., et al. (2002, July). *Students' use of modal clumps to summarize data*. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.
- Konold, C., & Miller, C. D. (2011). *Tinkerplots: Dynamic data exploration*. Emeryville, CA: Key Curriculum Press.
- Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, 41(3), 645–679.
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82–105.
- Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13(1–2), 152–173.
- Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Social Sciences*, 1(2). Retrieved from <http://www.qualitative-research.net/index.php/fqs/issue/view/28>
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- 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.
- Prediger, S., Link, M., Hinz, R., Hußmann, S., Thiele, J., & Ralle, B. (2012). Lehr-Lernprozesse initiieren und erforschen—fachdidaktische Entwicklungsforschung im Dortmunder Modell [Initiating and investigating teaching-learning processes—topic-specific didactical design research in the Dortmund model]. *Mathematischer und Naturwissenschaftlicher Unterricht*, 65(8), 452–457.
- Prediger, S., & Zwetzschler, L. (2013). Topic-specific design research with a focus on learning processes: The case of understanding algebraic equivalence in grade 8. In T. Plomp & N. Nieveen (Eds.), *Educational design research—Part A: An introduction* (pp. 409–423). Enschede, The Netherlands: SLO.

- Schnell, S., & Büscher, C. (2015). Individual concepts of students Comparing distribution. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 754–760).
- Stroeve, J. & Shuman, C. (2004). *Historical Arctic and Antarctic surface observational data, version 1*. Retrieved from <http://nsidc.org/data/nsidc-0190>.
- Vergnaud, G. (1990). Epistemology and psychology of mathematics education. In P. Neshier (Ed.), *ICMI study series. Mathematics and cognition: A research synthesis by the International Group for the Psychology of Mathematics Education* (pp. 14–30). Cambridge: Cambridge University Press.
- Vergnaud, G. (1996). The theory of conceptual fields. In L. P. Steffe (Ed.), *Theories of mathematical learning* (pp. 219–239). Mahwah, NJ: Lawrence Erlbaum Associates.

Chapter 3

Students Reasoning About Variation in Risk Context



José Antonio Orta Amaro and Ernesto A. Sánchez

Abstract This chapter explores students' reasoning about variation when they compare groups and have to interpret spread in terms of risk. In particular, we analyze the responses to two problems administered to 87 ninth-grade students. The first problem consists of losses and winnings in a hypothetical game; the second is about life expectancy of patients after medical treatments. The problems consist of comparing groups of data, and choosing one in which it is more advantageous to bet or to receive medical treatment. In this research we propose three levels of students' reasoning when they interpret variation. Decision making in the third level of reasoning is influenced by risk. As a conclusion, some characteristics of the problems and the solutions provided by the students are highlighted.

Keywords Middle school students · Reasoning · Risk · Statistics education
Variation

3.1 Introduction

In this study we analyze the responses provided by middle-school students (ages 14–16) to two elementary problems regarding comparison of data sets. In the problem design, we sought to promote the consideration of spread in comparing two data sets by proposing data sets with equal means and in a risk context. The analysis of the responses consists in identifying and characterizing the students' reasoning levels when they face such problems in order to understand how their reasoning on spread can improve. Spread refers to the statistical variation for data sets and

J. A. O. Amaro (✉)

Escuela Nacional para Maestras de Jardines de Niños, Mexico City, Mexico
e-mail: jaortaa@gmail.com

E. A. Sánchez

Departamento de Matemática Educativa, Centro de Investigación y de Estudios
Avanzados del Instituto Politécnico Nacional, Mexico City, Mexico
e-mail: esanchez@cinvestav.mx

© Springer Nature Switzerland AG 2019

G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_3

is one of the seven fundamental concepts of statistics (Burrill and Biehler 2011). As Watson (2006, p. 217) indicates, variation is the reason why statistics exists because it is ubiquitous and therefore, present in data. Garfield and Ben-Zvi (2008) observed that “Understanding the ideas of spread or variability of data [...] is essential for making statistical inferences” (p. 203). However, perceiving and understanding variability comprises a wide range of ideas. For instance, there is variability in data, samples, distributions and comparisons of data sets (Ben-Zvi 2004; Ben-Zvi and Garfield 2004; Shaughnessy 2007). This work is focused on the role of spread in the comparison of data sets.

In general, statistics problems comparing data sets involve deciding whether two or more sets can be considered equivalent or not. One way of doing so is through the comparison of center, spread and shape (Ciancetta 2007). Nevertheless, talking about the equivalence of data sets is difficult in basic school problems (Garfield and Ben-Zvi 2005). Therefore, among the problems used in research are those in which the subject is presented with two or more data sets and has to identify the one that has the highest quantity or intensity of the characteristic to be measured (e.g., grades, money, life expectancy). The statistical procedure to solve this type of tasks is based on the calculation of the mean of each set and the subsequent comparison of means. Regardless of its apparent simplicity, finding and reasoned using of this procedure represents a real difficulty for students in basic levels, who are inclined to other strategies (some of which are merely visual while others are based on isolated data from each set) instead of using the mean (Gal et al. 1989; Watson and Moritz 1999). After the study by Watson and Moritz (1999) there has been an increased interest to integrate the role of variation in the students’ analysis when they solve problems about data sets comparison. In this study, we carry on to pose a question with the same objective, proposing new problems to explore the students’ reasoning. We have particularly chosen problems in a risk context to evidence the uncertainty that spread usually uncovers. We then ask: How do students consider data spread in problems involving comparing data sets in a risk context?

3.2 Background

Gal et al. (1989) studied the intuitions and strategies of elementary-school students (3rd and 6th grades) when facing tasks of comparing data sets. The tasks were presented in two contexts: outcomes of frog jumping contests and scores on a school test. Several tasks were formulated per context and in each of them, two data sets were presented in a graph similar to those in Fig. 3.1. The students were asked whether both groups performed well or whether one did it better. Characteristics such as number of data, shape, center and spread of each couple of data sets were systematically manipulated to observe their influence on the students’ reasoning. The responses were divided into three categories: *statistical methods*, *proto-statistical methods*, and *other/task-specific methods*. *Statistical methods* included responses in which the sets were compared through data summaries for each set, particularly when arithmetic

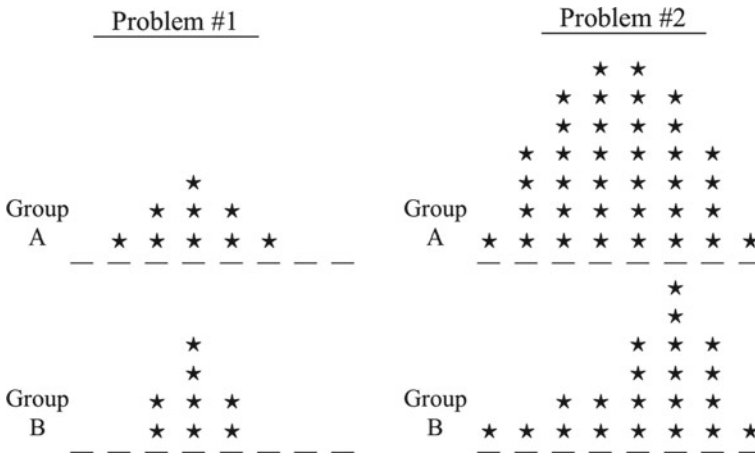


Fig. 3.1 Two of Gal et al.'s (1989) data sets comparison tasks. The contexts of ‘distances jumped by frogs’ and ‘class test scores’ were used for problems 1 and 2. For each problem, students compared groups A and B then decided if the groups did equally well or if one group did better. Taken from Ciancetta (2007)

means were used. The students whose responses were classified as *proto-statistical* ignored relevant characteristics of the data or did not summarize the information for each set; for example, they only compared modes. Those responses in which the students only added the data or provided qualitative arguments, as inferring that the team with the smaller number of frogs was better because they try harder, were classified as *other/task-specific methods*.

Watson and Moritz (1999) explored the structure of students’ thinking when they solve data sets comparison tasks. They adapted the protocol and four tasks by Gal et al. (1989) but only in the score context. In addition, they used the Structure of Observed Learning Outcomes (SOLO), a neoPiagetian model of cognitive functioning (Biggs and Collis 1982, 1991), to describe the levels of students’ responses, according to their structural complexity. The authors considered visual and numerical strategies and differentiated between “groups of equal size” and “groups of unequal size” to build a hierarchy of two cycles of UMR (Unistructural-Multistructural-Relational):

- U1: A single feature of the graph was used in simple group comparisons.
- M1: Multiple-step visual comparisons or numerical calculations were performed in sequence on absolute values for simple group comparisons.
- R1: All available information was integrated for a complete response for simple group comparisons; appropriate conclusions were restricted to comparisons with groups of equal size.
- U2: A single visual comparison was used appropriately in comparing groups of unequal sample size.
- M2: Multiple-step visual comparisons or numerical calculations were performed in sequence on a proportional basis to compare groups.

R2: All available information, from both visual comparison and calculation of means, was integrated to support a response in comparing groups of unequal sample size. (Watson and Moritz 1999, p. 158)

The differentiation between problems for data sets with the same size and unequal size is related to the use of means to compare. Still, students' consideration of spread was not involved in building the hierarchy. Three years later, Watson (2001) carried out another study, exploring the reasoning of students who had been interviewed in the first research (Watson and Moritz 1999) three years before. In this longitudinal study, besides formulating the research questions of the previous study, Watson posed the following question: "What evidence is shown that variation displayed in the data sets is explicitly considered in making decisions about which group did better?" (p. 343). The students' responses were clustered into six categories: (1) *No acknowledgment of variation*. (2) *Individual features—single columns* [of data]. (3) *Individual features—multiple columns* [of data]. (4) *Global features—'more'* [assumed to be based on visual comparisons]. (5) *Global features—multiple features*. (6) *Global features—integrated, compared and contrasted*. Watson (2002) deepened the same study using a new method, exploring whether the students' responses improve when there is an intervention in which a cognitive conflict in the subjects is created. Each student was shown a video of an interview with another student whose ideas were different from those of the student watching the video, in an attempt to make the student reflect and consider the possibility of changing his or her own conceptions.

Shaughnessy et al. (2004) conducted research to develop and study the conception of variability in middle and high-school students. The students answered a survey, and instructional interventions on sample distribution were carried out; finally, a sub-group of students was interviewed. Particularly, in the second interview, the researchers used the movie waiting time task shown in Fig. 3.2.

The responses to this task were coded in six categories: *Specific Data Points*, *Context*, *Centers*, *Variation*, *Distribution*, and *Informal Inferences*. Such categories were not mutually exclusive, so one response could be coded in two or more categories. When a response was based on the comparison of isolated points of each distribution, it was coded in *Specific Data Points*. When it referred to the student's personal experiences, it was classified in *Context*. Those responses using the means or medians were coded in *Centers* while those including comparisons of variation relative quantities (as considering ranges) were coded in *Variation*. If considerations on center and variation were combined, the response was coded in *Distribution*. Finally, if the response speculated on the probability of having certain experience in the waiting times at each movie theater and the subject used language including terms as 'predictable,' 'consistent,' 'reliable,' 'chances' or 'luck', it was coded in *Informal Inferences*. As a result, they found that most of the interviewed students considered both center and variation in their responses. Two-thirds of the interviewees stated that the two data sets were different from each other despite having the same mean and median. About 70% of them chose the movie theater with the least variation (Royal Theater). Nearly a third of the students included personal experiences in the responses.

Movie Waiting Time. A recent trend in movie theaters is to show commercials along with previews before the movie begins. The *wait-time* for a movie is the difference between the advertised start time (like in the paper) and the ACTUAL start time for the movie. A class of 21 students investigates the wait-times at two popular movie theater chains: Maximum Movie Theaters and Royal Movie Theaters. Each student attended two movies, a different movie in each theater, and recorded the wait-times in minutes below.

| | |
|--|---|
| <p>Maximum Movie Theaters:</p> <p>5.0 12.0 13.0 5.5 9.5 13.0 5.5</p> <p>11.5 8.0 8.5 14.0 13.0 8.5 7.0</p> <p>8.5 12.5 13.5 11.5 9.0 10.0 11.0</p> <p>Mean=10 minutes; Median=10 minutes</p> | <p>Royal Movie Theaters:</p> <p>11.5 11.0 9.0 10.5 8.5 11.0 9.0</p> <p>10.5 9.5 8.5 10.0 11.5 10.5 8.5</p> <p>9.0 11.0 11.0 9.5 10.0 9.0 11.0</p> <p>Mean = 10 minutes; Median = 10 minutes</p> |
|--|---|

Wait-Time for Movies

Maximum Movie Theaters

| | | | | | | | | | | | | | | |
|---|---|---|---|---|----|----|----|----|----|---|---|---|---|---|
| | | | | X | | | | | | X | | | | |
| | X | | | X | | | | | X | X | | | | |
| X | X | | X | X | X | X | X | X | X | X | X | X | X | X |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | | | | | |

Minutes (rounded to the nearest half-minute)

Royal Movie Theaters

| | | | | | | | | | | | | | | |
|---|---|---|---|---|----|----|----|----|----|--|--|--|--|--|
| | | | | X | | | X | | | | | | | |
| | | | | X | | | X | | | | | | | |
| | | | X | X | | X | X | | | | | | | |
| | | | X | X | X | X | X | X | | | | | | |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | | | | | |

Minutes (rounded to the nearest half-minute)

- What can you conclude about the wait-times for the two theaters?
- One student in the class argues that there is really no difference in wait-times for movies in both theaters, since the averages are the same. Do you agree or disagree? Why?
- Which of these theater chains would you choose to see a movie in? Why?

Fig. 3.2 Movie wait time task (Shaughnessy et al. 2004)

Orta and Sánchez (2011) explored how the notions of mean, range and uncertainty influenced the understanding of the statistical variability among middle-school students. The problem with which they collected the information was an adaptation of the Movie Wait Time Task by Shaughnessy et al. (2004). When student participants in the study of Orta and Sánchez (2011) were asked which movie theaters they would choose to watch a film at, and given that the three theaters were at the same distance from home, most of the justifications were based on personal experiences, without taking the data into account. The students’ responses included: “it is nearer home”, “I like those cinemas and I don’t mind watching the trailers they show” and “that

cinema is more famous”. Although the justifications based on personal experiences can be reasonable, in this study they were not based on data.

3.3 Conceptual Framework

The statistics education community has distinguished three overlapping areas of statistics to organize and analyze the objectives, activities and results of statistical learning: statistical literacy, reasoning, and thinking (Garfield and Ben-Zvi 2008). This study is located in the area of statistical reasoning. The purpose of the research on statistical reasoning is to understand how people reason with statistical ideas in order to propose features to create learning scenarios. When students try to justify their responses, elements that they think are important to the situation are revealed; in particular, the data they choose, operations made with these data and knowledge and beliefs on which they are based, are important in reasoning.

3.3.1 *Uncertainty and Decision Making*

Statistics is a general method to solve problems in situations where the subject deals with data, variation and uncertainty (Moore 1990). Particularly, Tal (2001) proposes that statistical variation that cannot be explained is uncertainty. We ask: Could considering variation as uncertainty in situations of data set comparison help in decision making? Before answering, we should observe that the contexts from which data arise promote the possibility of associating variation with uncertainty to a greater or lesser degree. For instance, in the context of jumping frogs and scores from Gal et al.’s research (1989), consider problem 1 of Fig. 3.1. It is not easy to think about the variation within each data set as an indicator of the uncertainty. In addition, choosing one or the other option carries no consequences to the person making the decision. Contrastingly, in the movie waiting time task by Shaughnessy et al. (2004), choosing the movie theater whose waiting time data have lesser variation means accepting a lesser uncertainty regarding the start time of the film. Even though the choice, in this case, has consequences for the person making the decisions, the students do not mind the uncertainty and can wait for some time at the movie theater without being affected. They value other characteristics more: closeness of the Movie Theater or comfort of the seats (Orta and Sánchez 2011). To promote the students’ perception of uncertainty and that their consideration will have consequences, it is convenient to choose a context in which the variability of the data involved is more directly related to uncertainty. It is also convenient that choosing one or the other set will have significant consequences. This could be achieved by using situations involving gaining or losing something valuable to the subject, such as health or money.

3.3.2 *Tasks in Risk Context*

At first glance, the notion of risk is related to an adverse event that may or may not occur. Aven and Renn (2010) suggest risk is related to an expected value, a distribution of probability as uncertainty or an event. According to Fischhoff and Kadwany (2011), risk is present when there are unwanted potential outcomes that may lead to losses or damages. To make problems involving data set comparison have consequences the subject considers relevant, a risk context seems promising. A paradigmatic task in risk context consists in making a decision about two games where gains and losses are at stake. Consider the following problem:

The gains of realizations of n times the game A and m times the game B are: Game A: $x_1, x_2 \dots; x_n$ Game B: $y_1, y_2 \dots; y_m$. Which of the two games would you choose to play?

The solution is reached using a flow diagram: (1) Compare \bar{x} and \bar{y} , (2) if $\bar{x} \neq \bar{y}$, then choose the game with the greatest mean; (3) if $\bar{x} = \bar{y}$, then there are two options: (3a) choose any game, or (3b) analyze the dispersion of data in each game and choose according to risk preferences. Two concepts of the theory of decision on risk (Kahneman and Tversky 1984) characterize risk preferences. Let us say that a preference is motivated by risk aversion when an option with data that have less spread over another with data that have greater spread is preferred. The decision is motivated by risk seeking when the option with data that have greater spread is chosen. For example, in their study Kahneman and Tversky (1984) proposed the participants to choose between 50% of probability of winning \$1000 and a 50% chance of not winning anything and the alternative of getting \$450 with certainty. Many subjects made a decision motivated by risk aversion since they prefer the certain winning, even though the first alternative has a higher expected value.

3.4 Method

3.4.1 *Participants*

The participants were 87 students (aged 14–16) from two different ninth-grade classes in a private school in Mexico City (last year of middle school). The Mexican middle-school students (7th to 9th grade) study data analysis and graphical representations. They deal with different statistical ideas such as arithmetic mean, range and mean deviation. In addition, they make, read and interpret graphics like bar graphs or histograms (SEP 2011). That is why we expected the students to make use of some of those statistical ideas to explain their answers, and most importantly, to use them in the context (risk). However, the actual responses of students did not meet our expectations.

3.4.2 Questionnaire

Two problems (which are presented below) were designed to explore the students' reasoning. Both problems in this research were designed by the authors to study the students' ideas about variation in situations in which "risk or uncertainty" were relevant. The authors considered different contributions of the research in the design. We first used the movie waiting time problem (Shaughnessy et al. 2004) to structure the problems in the questionnaire. When reading and analyzing the movie waiting time problem, we observe it deals with data set comparison. In addition, it has particular characteristics, such as, same number of cases, equal mean and median, and different spreads and bimodalities in both distributions. Ciancetta (2007, p. 103) considers that these qualities were included to promote the reasoning of variation in comparisons. From the structure of this problem (same number of data, equal mean and median, different spreads), we identified these as adequate situations to contextualize the problems of this research.

In the design of the gambling problem, we considered the work by Kahneman and Tversky (1984) in which the authors discuss that an analysis of decision making commonly differentiates between risk and riskless choices. Their paradigmatic example of decisions under risk is the acceptability of a game that leads to monetary results with specific probabilities. In the configuration of this problem, we also considered the idea by Bateman et al. (2007) regarding the introduction of a "small" loss as part of the game. This makes the game seem more attractive and promotes the students' reflection on the situation. Considering the problem of movie waiting time and the gambling situations as reference, we structured Problem 1 of the questionnaire as follows.

Problem 1. In a fair, the attendees are invited to participate in one of two games but not in both. In order to know which game to play, John observes, takes note and sorts the results of 10 people playing each game. The cash losses (−) or prizes (+) obtained by the 20 people are shown in the following lists:

Game 1: 15, −21, −4, 50, −2, 11, 13, −25, 16, −4

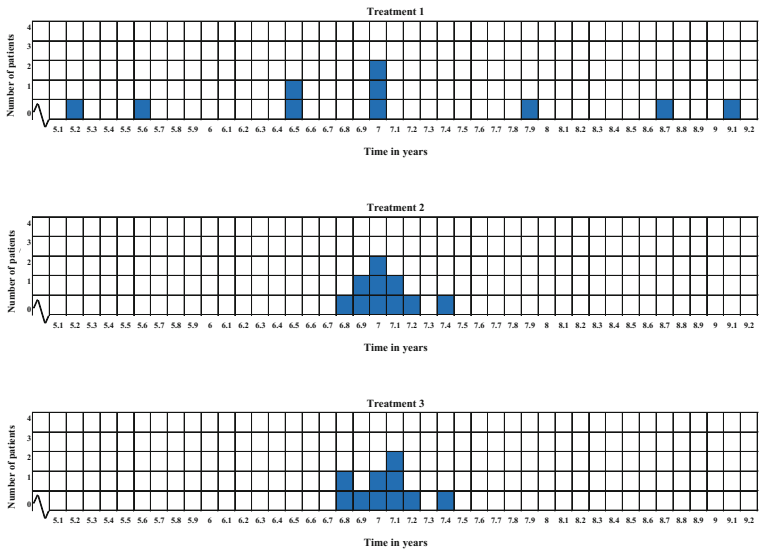
Game 2: 120, −120, 60, −24, −21, 133, −81, 96, −132, 18

- (a) If you could play only one of the two games, which one would you choose?
- (b) Why?

To create the problem regarding medical treatments, we considered the situation proposed in the research of Eraker and Sox (1981) on scenarios for palliative effects of medication for severe chronic diseases. In this situation, the authors present the choice between drugs that can extend life for several years. With this scenario and the structure of the movie waiting time problem, we created Problem 2 of the questionnaire as shown below.

Problem 2. Consider you must advise a person who suffers from a severe, incurable and deathly illness, which may be treated with a drug that may extend the patient's life

| | | | | | | | | | | |
|--------------|------|------|------|------|------|------|------|------|------|-----|
| Treatment 1: | 5.2, | 5.6, | 6.5, | 6.5, | 7.0, | 7.0, | 7.0, | 7.8, | 8.7, | 9.1 |
| Treatment 2: | 6.8, | 6.9, | 6.9, | 7.0, | 7.0, | 7.0, | 7.1, | 7.1, | 7.2, | 7.4 |
| Treatment 3: | 6.8, | 6.8, | 6.9, | 7.0, | 7.0, | 7.1, | 7.1, | 7.1, | 7.2, | 7.4 |



- a) Which treatment would you prefer (1, 2 or 3)?
- b) Why?

Fig. 3.3 Histograms of the three treatments

for several years. It is possible to choose between three different treatments. People show different effects to the medication: while in some cases the drugs have the desired results, in some others the effects may be more favorable or more adverse. The following lists show the number of years ten patients in each treatment have lived after being treated with one of the different options. Each number in the list corresponds to the time in years a patient has survived with the respective treatment. The graphs corresponding to the treatments are shown after (Fig. 3.3).

Table 3.1 shows the statistics of the Problems 1 and 2. Many of the characteristics are the same in both problems.

The general expected solution for each problem is discussed above in the conceptual framework. The problems above were solved in 50 min, approximately.

Table 3.1 Statistics of Problems 1 and 2

| Statistic | Problem 1 games | | Problem 2 treatments | | |
|-----------|-----------------|--------|----------------------|-------------|-------------|
| | Game 1 | Game 2 | Treatment 1 | Treatment 2 | Treatment 3 |
| Mean | 4.9 | 4.9 | 7.04 | 7.04 | 7.04 |
| Std. dev. | 21.43 | 96.83 | 1.23 | 0.17 | 0.18 |
| Range | 75 | 265 | 3.9 | 0.6 | 0.6 |
| Sum | 49 | 49 | 70.4 | 70.4 | 70.4 |
| Min | -25 | -132 | 5.2 | 6.8 | 6.8 |
| Max | 50 | 133 | 9.1 | 7.4 | 7.4 |
| Count | 10 | 10 | 10 | 10 | 10 |

3.4.3 Analysis Procedures

To analyze the responses of students, the chosen group for each of them was first observed; secondly, the responses were categorized in accordance with the strategies of comparison deduced from their justification. For that purpose, we followed the suggestions of Birks and Mills (2011) about identifying important words and groups of words in data for categorizing them and propose levels of reasoning.

The responses were organized to show different levels of the students' reasoning associated with variation. The first shows responses in which variability is not perceived. In addition, few relevant strategies are included to choose one group or another. In the second level, strategies can be considered relevant to choosing between sets of data but ignoring the variability in the data. In the third level, responses show perception of variability and a relevant strategy to choose between one set and the other.

3.5 Results

The students' responses were organized in three reasoning levels, considering the type of justification or explanation of the decision or preference made.

Level 1 groups the responses with circular or idiosyncratic arguments. The first are statements that consider there is a greater gain in the chosen game (Problem 1) or that the treatment allows living longer (Problem 2), but without including data in the argumentation; the second introduces beliefs or personal experiences.

Level 2 contains responses with justifications that include the explicit considerations of some or all the data in each set. In Problem 1, all the responses at this level obtain the totals and compare them. In contrast, most of the responses to Problem 2 compare isolated data from each set.

Level 3 is constituted by responses with argumentation combining and comparing more than 1 datum from each set. In these responses, we perceive the differences

Table 3.2 Responses to Problems 1 and 2 by reasoning level

| | Level 1 | Level 2 | Level 3 | No response | Total |
|-----------|----------|----------|----------|-------------|-----------|
| Problem 1 | 64 (74%) | 9 (10%) | 13 (15%) | 1 (1%) | 87 (100%) |
| Problem 2 | 49 (56%) | 25 (29%) | 13 (15%) | | 87 (100%) |

between the data sets in terms of the risk each game or treatment involves. Decision making in these cases is influenced by risk aversion or seeking.

Table 3.2 shows a number of responses to the problems in this research that were categorized in each level.

Below, we show examples of the three levels for each problem solved by the students.

3.5.1 Problem 1: Levels of Reasoning

Level 1. Circular argumentation. In this level, students made a choice but did not justify it based on any treatment of the data. At first glance, the data seemed to suggest that there was more to win or lose (60 responses) in the games. Regardless, they did not specify this and only provided circular arguments such as “because you win more than in 1” (Fig. 3.4). When looking at the data, the students were likely to focus only on some of them and mentally compared them (attention bias). Some characteristics in the responses allowed us to deduce that the students compared specific data. For example: they compared one or two of the greatest data values in a set with one or two from the other. Some focused on the highest losses in each set while others paid attention only to the number of data values with positive (or negative) signs and compared them with the corresponding number of the other set (4 responses).

While in the hierarchy by Watson there are levels that consider numerical and visual strategies, this are surely motivated by the graphic presentation of the data set and detected by the researchers thanks to the interviews. In the case of our study, none of these conditions were presented: the presentation of the data did not include graphs and we conducted no interviews. Visual strategies are those that arise from graphical data observation; however, this research did not include them. However, as stated before, most of the strategies in this level were based on observing, at first

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?

El 2
 ¿Por qué? porque ganas mas que en el 1

Fig. 3.4 Example of Level 1 response to Problem 1 (“because you win more than in 1”)

| | | | | | | | | | | |
|----------|-----|------|------|-----|-----|-----|-----|-----|------|---------|
| Juego 1: | | -56 | | | | | | | | |
| | 15 | -21 | -4 | 50 | -2 | 11 | 13 | -25 | 16 | -4 (49) |
| Juego 2: | | | -378 | | | | | | | |
| | 120 | -120 | 60 | -24 | -21 | 133 | -81 | 96 | -132 | 18 (49) |

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?

cualequiera.
 ¿Por qué? terminar ganando lo mismo.

Fig. 3.5 Example of Level 2 response to Problem 1 (“you end up winning the same”)

glance, one or two elements in each set and compared each other. In most of the cases, we were unable to determine which specific values students observed in each set because their arguments were circular and there were no interviews afterwards to clarify them. As for the hierarchy by Shaughnessy et al. (2004), Level 1 is similar to the “No acknowledgement of variation” level and partially similar to the “Individual features” level.

Level 2. Data consideration: Totals. In this reasoning level, students summed the data in each set and compared the totals (nine responses). From this, students usually argued that it was possible to choose any game because “you end up winning the same” (Fig. 3.5). This strategy enclosed the origin of the statistical procedure of combining observations. For the game problem we analyzed here, the totals are adequate numbers to make the comparison. In addition, we considered all the data and the results were not evident at first glance, but they demanded a certain treatment. The fact that both sets had the same number of data did not allow for differentiating among the students who would perceive the importance of the size of the data sets (proportionality) from those who would not. In the responses placed in this level data variability was ignored or not acknowledged, and risk was not detected.

This reasoning level would be included in Watson’s level M1 because it includes numerical calculations to compare two sets with the same number of data. However, it is not comparable to any level in the hierarchy proposed by Shaughnessy et al. (2004) since mean and median are part of the information given in the data presentation of the problem regarding movie waiting time. Therefore, students did not have to sum data nor obtain the mean, although the reference to mean would appear in some students’ considerations.

Level 3. Data combination: Risk. In this reasoning level, we included the responses that provided characteristics to indicate that students perceived risk. In general, the strategies reflected in the responses included in this level consist of simultaneously focusing attention on the relationship between what can be won in each game (maximums) and what can be lost (minimums). The consideration of

a) Si tienes la posibilidad de participar en un solo juego ¿Cuál juego elegirías?

el juego 2
 ¿Por qué? porque hay más posibilidad de ganar más, aunque también se pierde más dinero.

Fig. 3.6 Example of a Level 3 response to Problem 1 (“because there is a bigger possibility of winning more, although you also lose more money”)

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)?

1
 ¿Por qué? se pueden vivir más años que con otros tratamientos

Fig. 3.7 Example of Level 1 response to Problem 2 (“one can live for more years than with other treatments”)

choosing game 2 “because there is a bigger possibility of winning more, although you also lose more money” (Fig. 3.6), led the students to notice that the games were not equivalent and risk in one or the other differentiates the games. The choice is not entirely determined by the student’s analysis, but also by their risk preferences. For example, a student chose game 1 as the most convenient and justified the choice by stating “Because as game 2, [game 1] has losses, but in a lower number and you don’t risk that much”. The student’s choice was influenced by *risk aversion*. Interestingly, all the responses, with the exception of one, did not mention the equality of the totals or means of the data sets. This did not affect the analysis that they were equal in both sets. In the only response that considered both the totals and the spread, the student chose any game “because you have the same chance of winning or losing; in one you don’t win or lose much, in the other you win a lot or lose a lot”. In the characteristics written, we perceive that the student summed and obtained 49 in both data sets. Despite noticing that one game was riskier than the other, the student did not prefer any.

3.5.2 Problem 2: Reasoning Levels

Level 1. Circular argumentation. In this level we included the responses in which a treatment was chosen and the preference was only justified with expressions such as “one lives longer” (32 responses were coded with this argument), or by stating the choice “looking at the graph” (five responses) or providing idiosyncratic justifications (12 responses). In any case, those who said “one can live for more years than with other treatments” (Fig. 3.7) did not clarify why one would live longer with one treatment or with the other. Some other responses did not specify which part of the graph they considered relevant. In the responses coded as idiosyncratic, they introduced personal beliefs that were not relevant to the problem.

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? el 2
 ¿Por qué? porque ahí me aseguran por lo menos 7 años de vida más

Fig. 3.8 Example of Level 2 response to Problem 2 (“because you ensure at least 7 years more of life”)

In this problem of medical treatments, the graphical representation of the data promoted the use of visual strategy among the students. However, they did not often explain what they saw in the graph. This level is similar to the level U1 in Watson’s hierarchy, with the difference that she obtained the data through interviews, so that the students could reveal their strategy.

Level 2. Data consideration. In this level we classified the responses in which there was argumentation to favor one treatment based on specific values of each set and their comparison (23 responses). We also included the responses in which the data from each set were summed, and the totals were compared (two responses). From the 23 responses above, eight were based on the observation of the maximums and the minimums. For example, a student chose treatment 1 “because there is a greater chance of living for 9 years”. In 15 responses, the modal value was mentioned: a student chose treatment 2 “because you ensure at least 7 years more of life” (Fig. 3.8), but we did not know how the modal value was used to make the decision. We supposed the student combined it with the observation of one or two extremes without stating so.

Level 3. Data combination: Risk. In this level, we grouped the responses based on the consideration of two or more data values that allowed students to perceive variability. In 11 responses, the choice was justified mentioning both the maximum and the minimum. For example, a student chose treatment 2 “because I will probably won’t live 9 years more, but I can ensure from 6.8 to 7.4 years.” Two responses mentioned the mode and some extreme; these students chose treatment 3 because the mode was higher than in the other two, but they also considered some extreme values. As an example, a student chose treatment 3 “because in the graphs the minimum is 6.8 years and the most frequent [period] is 7.1 while in treatment 1, 5.2 is the minimum and 7 is the most frequent.” In several of these responses, the students provided characteristics that allowed us to suppose that they perceived risk; for example; when they stated “because I may not live 9 years more, but I still have from 6.8 to 7.4 for sure” (Fig. 3.9) and chose treatments 2 and 3, avoiding the risk of living only 5.2 years, as with treatment 1. We can suppose that their choice was influenced by risk aversion. On the other hand, the responses in which they chose treatment 1, and that considered that they “can live up to 9 years although [one] could live only for 5.2 years”, were influenced by risk seeking.

a) ¿Qué tipo de tratamiento preferirías (1, 2 ó 3)? Tratamiento 03
 ¿Por qué? Porque tal vez no viva 09 años más
pero así tengo asegurada de 6.8 a 7.4

Fig. 3.9 Example of Level 3 response to Problem 2 (“because I may not live 9 years more, but I still have from 6.8 to 7.4 for sure”)

3.6 Discussion

It is convenient to make some observations that may be useful to answer the question posed at the beginning of this work and clarify the purpose of this research. Particularly, we will consider some characteristics of the problem and the solutions provided by the students.

In relation with the characteristics of the problem, we emphasize that spread can be significantly interpreted to use it in decision making in risk contexts. In the context of winning in games, we easily perceive that the game with the highest spread is riskier because there is more to win, but there is also more to lose. In contrast, in the context of score problems by Gal et al. (1989) and Watson and Moritz (1999), it is unclear how the spread of data can allow making a decision on the group that performs better.

In data set comparison problems, when the sets to compare have the same number of data and same mean (and median) but have different spreads, it is unclear how spread is used to make a decision. For instance, Gal’s problem (1989) presents the data in Fig. 3.10; these data represent the score that students in Groups A and B obtained in an exam. The question posed is: Which group performed better?

In Group A, a student scored 7, better than any other in Group B; however, another student obtained a score of 3, which is worse than the score of any other student in Group B. There are three students in Group A whose score is the mean while there are five students in Group B in the same condition. Do these observations have implications that allow deciding which group is better? They apparently do not clearly help to decide if one group performs better than the other when no other criteria are added. If the consideration on variation has no implications, then why should it be done?

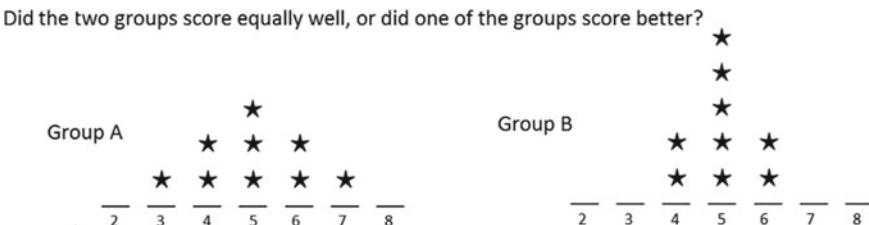


Fig. 3.10 One of the problems of Gal et al. (1989) adapted

We have observed that, when the data context is winning a game or life expectancy after a medical treatment, several students place themselves in position where they are benefitted or affected by one decision or the other. They realize that the game or treatment corresponding to the set with the highest spread offers them the possibility of winning more or living longer, but also of losing more or living for a shorter time. On the other hand, the game or treatment with the least spread data set offers them the least uncertainty. As in Gal's problem, there is no regulatory outcome to indicate which the best decision is because such response depends on the risk preference of the person making the decision. This is an uncertainty pattern frequently present in daily life: Do we prefer a stable job or business, even if we earn a little or one in which we earn more but carries greater risk? The concepts of risk aversion or seeking help us to understand how we decide when facing those options.

The reasoning levels that have been defined are characterized by the presence or absence of data in the justification as well as their combination. Given that what we analyze are the students' written responses not their thoughts, the levels reflect their ability to write the idea that leads them to make a decision. As mentioned before, the hierarchy we propose, is different from the ones by Watson and Moritz (1999) and Shaughnessy et al. (2004) in that they conducted interviews, hence the nature of their data is different.

In *Level 1*, the students' argumentation does not include the data they used to make the decision. Most of the students "justify" their decisions using circular arguments, that is, arguments that only claim what they should prove. This shows that many of them are not conscious that a key point of argumentation is to evidence of the way in which data were used to make a decision. They surely saw something in the data that affected their response, but they did not write that in their response.

In *Level 2*, the choice and the use of data to reach a conclusion are shown in the argument. However, either the selection is not the best to make a decision or the use of data is not the best. In this level we grouped the responses with arguments that do not consider variation. A number of them sum the data in each set and compare them. This strategy is partial because it allows the students to perceive the equality of the means because the sets have the same number of elements, but the students fail to consider the spread. Others only consider isolated data, as comparing the maximums, the minimums or the ranges. If they considered more than one of these data, they did not link them properly; therefore, they did not perceive spread.

Level 3 shows the choice and use of data in the argument to reach a conclusion. The data are also combined to provide a notion of spread and risk. The students did not use a measure of spread but coordinated more than one data value, mainly maximums and minimums, to perceive that the games in situation 1 and the treatments in situation 2 were different due to the risk they posed. In this way, the measure of spread they implicitly used was the range and with it they perceived the risk.

In the design of the present study we sought to reveal the students' reasoning regarding spread in data set comparison, and we did not try to reveal their reasoning about and with the notion of center, particularly the arithmetic mean. Regardless, the total absence of a strategy to calculate the arithmetic mean of each data set and the comparison between them is remarkable. The closest thing to this strategy is

calculating the total of each set and comparing them. Clearly, students do not think of these totals as representatives of each set but as the amount of money or time involved in them. Although technically, the only missing element is dividing between the total data to have a comparison based on the means, the conceptual distance between the idea behind the consideration of the totals and the idea of representative of a data set is quite large.

3.7 Conclusion, Limitations and Implications

Based on the results obtained in this research, the risk context provides an alternative with which students engage and in which the words or numbers used by them to solve the problems are numbers and words in context, a central idea in statistics. When solving the problems raised, the students associated risk and variability; the latter was described through the range of the data sets. As part of the analysis, the students used the sum of all the data available; however, they did not manage to complete the algorithm for the arithmetic mean. There is a void in the students' analyses; they do not use center and dispersion of a data set for making a decision. Research deriving from this should pose a learning trajectory that guides students to integrate the notions of center and dispersion. This would enable students to improve their analysis of data sets or distributions.

The presence of students' reasoning considering risk shows the convenience of looking for situations in this context to suggest problems related to dispersion. Together with other scenarios already used in teaching to promote learning and consideration of dispersion, the concept of risk might be another source of problem situations for statistics teaching at mid-school level. Accordingly, a deeper research might find other risk situations to explore students' reasoning and promote their development in the statistics class. Additionally, the reasoning levels students can reach when facing those problems and situations could be characterized with greater precision. Additionally, the reasoning levels students can reach when facing those problems and situations could be characterized with greater precision. The use of technology is a necessary complement for a more thorough research that allows for easier calculations and emphasis on conceptual discussions. Furthermore, technology provides options for simulations, which could correspond to distributions related to the game or medical treatment contexts.

The risk models that truly work in society are very complex, even though gains/losses situations are common in business. Decisions and consequences hardly derive only from such situations despite the fact that the calculation of probabilities, centers, and dispersion are frequently part of the analysis. This is a limitation in developing the option of promoting risk situations in statistics teaching at middle school. Still, we consider its possibilities have not been sufficiently explored for statistics teaching at other educational levels.

Acknowledgements Grant EDU2016-74848-P (FEDER, AEI). Grant CONACYT No. 254301.

References

- Aven, T., & Renn, O. (2010). *Risk, governance and society*. Berlin: Springer.
- Bateman, I., Dent, S., Peters, E., Slovic, P., & Starmer, C. (2007). The affect heuristic and the attractiveness of simple gambles. *Journal of Behavioral Decision Making*, 20(4), 365–380.
- 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 reasoning about variability. *Statistics Education Research Journal*, 3(2), 4–6.
- Biggs, J. B., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Biggs, J., & Collis, K. (1991). Multimodal learning and the quality of intelligent behaviour. In H. Rowe (Ed.), *Intelligence, reconceptualization and measurement* (pp. 57–76). Hillsdale, NJ: Laurence Erlbaum Associates.
- Birks, M., & Mills, J. (2011). *Grounded theory*. Thousand Oaks, CA: Sage.
- 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). New York: Springer.
- Ciancetta, M. (2007). *Statistics students reasoning when comparing distributions of data* (Doctoral thesis). Portland State University. Online: <http://www.stat.auckland.ac.nz/~iase/publications/dissertations/07.Ciancetta.Dissertation.pdf>.
- Eraker, S. E., & Sox, H. C. (1981). Assessment of patients' preferences for therapeutic outcomes. *Medical Decision Making*, 1(1), 29–39.
- Fischhoff, B., & Kadavy, J. (2011). *Risk: A very short introduction*. Oxford: Oxford University Press.
- 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, MO.
- 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 teaching practice*. New York: Springer.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39(4), 341–350.
- 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.
- Orta, A., & Sánchez, E. (2011). Influential aspects in middle school students' understanding of statistics variation. In M. Pytlak, T. Rowland & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society For Research in Mathematics Education, Rzeszów, Poland*.
- SEP. (2011). *Programas de Estudio. Educación Básica. Secundaria. Matemáticas*. México, D. F.: Secretaría de Educación Pública.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1010). Greenwich, CT: Information Age Publishing, Inc. and National Council of Teachers of Mathematics.
- Shaughnessy, J. M., Ciancetta, M., Best, K., & Canada, D. (2004). *Students' attention to variability when comparing distributions*. Paper Presented at the Research Pre-session of the 82nd Annual Meeting of the National Council of Teachers of Mathematics, Philadelphia, PA.
- Tal, J. (2001). *Reading between the numbers: Statistical thinking in everyday life*. New York: McGraw-Hill.
- Watson, J. M. (2001). Longitudinal development of inferential reasoning by school students. *Educational Studies in Mathematics*, 47(3), 337–372.

- Watson, J. M. (2002). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics*, 51(3), 225–256.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.
- Watson, J.M., & Moritz, J.B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37(2), 145–168.

Chapter 4

Students' Aggregate Reasoning with Covariation



Keren Aridor and Dani Ben-Zvi

Abstract Helping students interpret and evaluate the relations between two variables is challenging. This chapter examines how students' aggregate reasoning with covariation (ARwC) emerged while they modeled a real phenomenon and drew informal statistical inferences in an inquiry-based learning environment using TinkerPlots™. We focus in this illustrative case study on the emergent ARwC of two fifth-graders (aged 11) involved in statistical data analysis and modelling activities and in growing samples investigations. We elucidate four aspects of the students' articulations of ARwC as they explored the relations between two variables in a small real sample and constructed and improved a model of the predicted relations in the population. We finally discuss implications and limitations of the results. This article contributes to the study of young students' aggregate reasoning and the role of models in developing such reasoning.

Keywords Aggregate reasoning · Exploratory data analysis · Growing samples
Informal statistical inference · Reasoning with covariation · Statistical modelling

4.1 Introduction

The purpose of this chapter is to provide an initial scheme for understanding young students' emergent articulations of aggregate reasoning with covariation (ARwC) in the context of informal statistical inference from growing data samples. Handling data from an aggregate point of view is a core aspect of statistical reasoning (Hancock et al. 1992). Predicting properties of the aggregate is the essential aspect of data analysis and statistical inference. To achieve this goal, one should develop a notion of data as an organizing structure that enables seeing the data as a whole (Bakker and Hoffmann

K. Aridor (✉) · D. Ben-Zvi
Faculty of Education, The University of Haifa, Haifa, Israel
e-mail: kerenaridor@gmail.com

D. Ben-Zvi
e-mail: dbenzvi@univ.haifa.ac.il

© Springer Nature Switzerland AG 2019
G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_4

2005). The development of such aggregate view of data is a key challenge in statistics education (Bakker et al. 2004). Previous research about students' reasoning with covariation specified aspects of ARwC that are essential to judging and interpreting relations between two variables, such as viewing data as a whole (Moritz 2004). Nevertheless, young students tend to see data as individual cases (local view) rather than a global entity (Ben-Zvi and Arcavi 2001), and often focus on a single variable and not on the bivariate relationship (Zieffler and Garfield 2009). Statistical modelling contexts can help address these challenges by supporting students' search for patterns in data and accounting for variability in these patterns (Pfannkuch and Wild 2004).

To examine how young students' ARwC can emerge while they make informal statistical inferences and model an authentic phenomenon using hands-on tools and TinkerPlotsTM (Konold and Miller 2011), we first elaborate on the notions of informal statistical inference, covariational and aggregate reasoning. We then highlight reasoning with modelling and the "growing samples" pedagogy. In the Method section, we describe the Connections project and the fifth grade-learning trajectory to put the tasks of the project in context. Next, we present the main results of this research by specifying the four aspects that structure the fifth grade students' ARwC. We conclude with theoretical and pedagogical implications and limitations of the research.

4.2 Literature Review

4.2.1 Informal Statistical Inference (ISI)

Statistical inference moves beyond the data in hand to draw conclusions about some wider universe, taking into account uncertainty in these conclusions and the omnipresence of variability (Moore 2004). Informal Statistical Inference (ISI) is a theoretical and pedagogical approach for developing statistical reasoning, connecting between key statistical ideas and informal aspects of learning statistical inference (Garfield and Ben-Zvi 2008). ISI is based on generalizing beyond the given data, expressing uncertainty with a probabilistic language, and using data as evidence for these generalizations (Makar and Rubin 2009, 2017). The reasoning process leading to making ISIs is termed Informal Inferential Reasoning (IIR). IIR refers to the cognitive activities involved in informally formulating generalizations (e.g., conclusions, predictions) about "some wider universe" from random samples, using various statistical tools, while considering and articulating evidence and uncertainty (Makar et al. 2011). IIR includes reasoning with several key statistical ideas such as: sample size, sampling variability, controlling for bias, uncertainty and properties of data aggregates (Rubin et al. 2006).

4.2.2 Reasoning with Covariation

This chapter is focused on statistical reasoning with covariation. Statistical covariation relates to the correspondence of variation of two variables that vary along numerical scales (Moritz 2004). Bivariate relations are characterized by the variability of each of the variables; the pattern of a relation, the shape of the relationship in terms of linearity, clusters and outliers; and the existence, direction and strength of a trend (Watkins et al. 2004). Reasoning with covariation is defined as the cognitive activities involved in coordinating, explaining and generalizing two varying quantities while attending to the ways in which they change in relation to each other (Carlson et al. 2002).

Reasoning with covariation plays an important role in scientific reasoning and is applied depending on usage, goals and discipline (Schauble 1996). For example, covariation can serve as an alternative to the concept of function. A covariation approach in this context entails being able to move between values of one variable and coordinating this shift with movement between corresponding values of another variable. Such an approach plays an important role in students' understanding, representing and interpreting of the rate of change, and its properties in graphs (Carlson et al. 2002). The approach can also lead to reasoning about the algebraic representation of a function (Confrey and Smith 1994).

Moritz (2004) identified four levels of verbal and numerical graph interpretations while analysing bivariate associations: Nonstatistical, single aspect, inadequate covariation and appropriate covariation. Nonstatistical responses relate to the context or to a few data points, such as outliers or extreme values, without addressing covariation. Single aspects responses refer to a single data point or to one of the variables (usually the dependent), with no interpolating. Inadequate Covariation responses address both variables but either relate to correspondence by comparing two or more points without generalizing to the whole data or to the population; or, variables are described without relating to the correspondence or by mentioning it incorrectly. Appropriate covariation responses refer to both variables and their correspondence correctly.

Moritz's hierarchy as well as other studies reflect students' challenges while reasoning with covariation. Students tend to focus on isolated data points rather than on the global data set and trend; focus on a single variable rather than the bivariate data; expect a perfect correspondence between variables, without exception in data (a deterministic approach); consider a relation between variables only if it is positive (the unidirectional misconception) ; reject negative covariations when they are contradictory to their prior beliefs; have a hard time distinguishing between arbitrary and structural covariation (Batanero et al. 1997; Ben-Zvi and Arcavi 2001; Moritz 2004).

Several studies suggest that a meaningful context for reasoning with aggregate aspects of distribution, such as shape and variability, can support developing reasoning with covariation (e.g., Cobb et al. 2003; Konold 2002; Moritz 2004; Zieffler and Garfield 2009).

4.2.3 Aggregate Reasoning

Developing statistical reasoning involves flexibly shifting between a local view of data and a global view of data according to the need and the purpose of the investigation (Ben-Zvi and Arcavi 2001; Konold et al. 2015). Aggregate reasoning is a global view of data that attends to aggregate features of data sets and their propensities (Ben-Zvi and Arcavi 2001; Shaughnessy 2007). When viewing data as an aggregate, a data set is considered as an entity with emergent properties, which are different from the properties of the individual cases themselves (Friel 2007). Two important aggregate properties are the distinction between signal and noise and the recognition and diagnosis of various types and sources of variability (Rubin et al. 2006).

Aggregate reasoning is discussed in the literature mostly in the context of data and distribution. The notion of distribution as an organizing conceptual structure is conceived by aggregate aspects of distribution, such as the general shape, how spread out the cases are, and where the cases tend to be concentrated within the distribution (Bakker and Gravemeijer 2004; Konold et al. 2015). Reasoning with bivariate data is mostly discussed without using the terminology of aggregate reasoning. For example, Ben-Zvi and Arcavi (2001) describe the way students' previous knowledge and different types of local observations supported and hindered the development of their global view of data. In the beginning, they reasoned with the investigated association as an algebraic pattern, with relation to local data cases and adjacent difference. However, this focus on pointwise observation eventually supported the development of the students' reasoning with the notion of trend while relating to the data as a whole.

Konold (2002) recognizes the gap between people's ability to make reasonable judgments about relations in the real world and their struggle to make judgments about covariation from representations such as scatterplots and two-by-two contingency tables. Konold suggests that this struggle stems from a difficulty to decode the ways in which these relationships were displayed (Cobb et al. 2003; Konold 2002). One goal of the current study is to extend the understanding of aggregate reasoning to the context of statistical modelling and covariation, which we term ARwC. The analysis of ARwC will consider various aspects of students' aggregate reasoning including reasoning with variability.

4.2.4 Reasoning with Variability

Variability is the aptness or tendency of something to vary or change (Reading and Shaughnessy 2004). Variability is omnipresent in data, samples and distributions (Moore 2004). While reasoning with data, students should search for signals in the variability, as well as for potential sources of such variability (Shaughnessy 2007). A signal can be considered as the patterns which have not been discounted as ephemeral. Such patterns can become evident only in the aggregate. Noise can be considered as

the unexplained variability around these patterns, if identified (Wild and Pfannkuch 1999).

Reasoning with variability has both informal and formal aspects, from understanding that data vary, to understanding and interpreting formal measures of variability. Students seem to reason intuitively with informal aspects of variability, such as the representativeness of variability by spread and the idea that data vary. However, students tend to focus primarily on outliers and have difficulties measuring variability in a way that depicts thinking of variability as representing spread around the center (Garfield et al. 2007).

A conceptual understanding of variability includes: (a) developing intuitive ideas about variability (e.g., repeated measurement on the same characteristic are variable); (b) the ability to describe and represent variability (e.g., the role of different representation of a data set in revealing different aspects of variability, the representativeness of spread measurements); (c) using variability to make comparisons; (d) recognizing variability in special types of distributions (e.g., the role of the variability of both variables' distributions to a bivariate data distribution); (e) identifying patterns of variability in fitting models; (f) using variability to predict random samples or outcomes; and (g) considering variability as part of statistical thinking (Garfield and Ben-Zvi 2005). Modelling a phenomenon entails the search for differences and similarities in the population, which is an initial step toward reasoning with variability (Lehrer and Schauble 2012).

4.2.5 Statistical Models and Modelling

Freudenthal (1991) viewed mathematics as a human activity. As such, students should learn mathematics by “mathematising”: they should find their own levels of mathematics and explore the paths leading there with as much guidance as they need. The process of “mathematising” lasts as long as reality is changing and extending under various influences, including the influence of mathematics, which is absorbed by this changing reality. One component of “mathematising” is modelling. As such, modelling is defined as simplifying or grasping the essentials of a static or dynamic situation within a rich and dynamic context (Freudenthal 1991). Modelling can be perceived as interrelating processes in which the role of the model is changing as thinking progresses. At the first process, a model emerges as a “model of” informal reasoning and develops into a “model for” more formal reasoning. At the second process, a new view of a concept emerges along the transition from “model of” to “model for”. Such view can be perceived as formal in relation to the initial disposition toward this concept. These two processes are accompanied by a third one—the shaping of a model as a series of signs that specifies the previous reasoning process (Gravemeijer 1999).

Models and modelling are essential components of statistical reasoning and thinking (Wild and Pfannkuch 1999). The practice of statistics can be considered as a form of modelling, as the development of models of data, variability and chance are paving

the way for a statistical investigation (Lehrer and English 2017). The modelling process entails an evaluation and improvement of models to include new theoretical ideas or data based findings (Dvir and Ben-Zvi 2018; Lesh et al. 2002). A statistical model is an analogy that simplifies a real phenomenon, describes some of the connections and relations among its components, and attends to uncertainty (Wild and Pfannkuch 1999). Based on the recognition that aggregate reasoning requires summarizing and representing data in multiple ways depending on the nature of the data, various pedagogical approaches have been developed. With this in mind, a modelling pedagogical approach can support the emergence of aggregate views of data (Lehrer and Schauble 2004; Pfannkuch and Wild 2004). In this chapter, we focus on students' emergent ARwC in relation to statistical models that were developed by them to describe a real phenomenon and predict outcomes for an unknown population. These models were constructed as part of the Connections learning environment, which was built on the growing samples and the purpose and utility ideas.

4.2.6 Task Design

The *growing samples* educational approach is an instructional idea mentioned by Konold and Pollatsek (2002), worked out by Bakker (2004) and elaborated by others (e.g., Ben-Zvi et al. 2012). In this approach, students are introduced to increasing sample sizes that are taken from the same population. For each sample, they pose a research question, organize and interpret the data, and draw ISIs. Later, they face “what if” questions that encourage them to make conjectures about same sized samples, or about a larger sample. In this approach, students are required to search for and reason with aggregate features of distributions and to identify signals out of noise. They need to compare their conjectures about the larger samples with insights from the data, to account for the limitations of their inferences and to confront uncertainty with regard to their inferences. The growing samples approach can be a useful pedagogical tool to support coherent reasoning with key statistical ideas (Bakker 2004; Ben-Zvi et al. 2012).

Another task design approach used in the Connections project was *purpose and utility*. Although everyday contexts can support learning statistics or mathematics, the strength of meaningful learning environments is in the design for purpose and utility (Ben-Zvi et al. 2018). The term purpose refers to students' perceptions. A purposeful task is a task that has a meaningful outcome (a product or a solution) for students. Such a purpose might be different from the teacher's intentions. The utility of ideas means that the learning process involves construction of meaning for the ways in which these ideas are useful. Purpose and utility are strongly connected. Purposeful tasks, encompass opportunities for students to learn to use an idea in ways that allow them to reason with its utility, by applying it in that purposeful context (Ainley et al. 2006). With this literature review in mind, we now formulate the research question of this study.

4.3 Research Question

In this case study, we focus on two fifth grade students (age 11) who were involved in modelling activities of bivariate data and drawing ISIs in growing samples investigations. In this context, we ask: *What can be the characteristics of the students' emergent ARwC?*

4.4 Method

4.4.1 The Setting

To address this question we draw on data from the 2015 Connections Project in a fifth grade Israeli classroom. In this project, a group of researchers and teachers designed and studied an inquiry-based learning environment to develop statistical reasoning. The focus of the 2015 Project was aggregate reasoning using modelling activities in the context of making ISIs. The design of the learning trajectory was guided by three main approaches: growing samples, statistical modelling, and purpose and utility. Students investigated samples of increasing size that were drawn from the same population. The goal of each investigation was to model an authentic phenomenon within the target population—all fifth graders in Israel. To do that, the students reasoned with the meaning and utility of statistical concepts, such as data, center, variability and distributions, using hands-on tools (pen and paper) or TinkerPlots™ (Konold and Miller 2011). TinkerPlots™ is an innovative data analysis, visualization and modelling tool designed to support students' (grades four to nine) reasoning with data.¹ TinkerPlots™ provides a dynamic graph construction tool that allows students to invent their own elementary graphs and evaluate them (Biehler et al. 2013). Models were constructed using the TinkerPlots™ Sampler, which can be used to model probabilistic processes and to generate random data from a model.

The students participated in ten activities (28 lessons, 45 min each) organized in two main cycles of data investigations of samples (see Table 4.1): (1) their whole class and grade (2–3 attributes, samples of 25 and 73 cases); and (2) their grade (18 attributes, samples of 10, 24 and 62 cases). For each sample, students posed a research question, organized and interpreted the data, and drew ISIs. They made conjectures about a larger sample to confront uncertainty. They modeled their conjecture about the investigated phenomenon in the target population, first as hands-on and after a while as TinkerPlots™ representations. Handouts included questions, such as, “would your inference apply also to a larger group of students such as the whole class?” Each lesson of the first nine activities included a whole class introductory discussion of the investigated topic, data investigation in small groups, and a whole class synthesis discussion of students' findings. In the tenth activity students summarized

¹www.tinkerplots.com

their investigations to present them in a student-parents event and write a final report about their findings. To build a collaborative culture of inquiry in the classroom, we encouraged the students to share ideas, products and actions, reflect about the learning processes and share their insights. The intervention began with a statistics pre-test and ended with an identical post-test that focused on reasoning with data, distribution, covariation, and informal inference. This learning trajectory preceded and significantly expanded the instruction according to the national fifth grade statistics curriculum, which focuses only on mean and median in a procedural manner.

4.4.2 The Participants

We fully documented the learning processes of 12 pairs of students. In this study, we focus on the development of ARwC of a pair of boys—Orr and Guy. Orr is an academically successful student who has high achievements in mathematics and science. He has a learning disability, which limits somewhat his ability to express himself verbally or in writing. Guy's academic achievements are usually low. The students were selected due to high motivation, creativity and interest in the investigation.

In the first six activities of the learning trajectory (Table 4.1), Orr and Guy investigated univariate distributions and associations between categorical and numerical attributes. During whole class meetings, they discussed statistical ideas, such as center and representativeness, variability and outliers, comparing groups and covariation, and ways to represent and articulate these ideas. During the seventh to the ninth activities, the data investigations dealt with the relations between the amount of push-ups one can make in a row ("Push-ups"), and the 900-m running time in seconds ("Running").

4.4.3 Data Collection and Analysis

The students' investigations were fully videotaped using Camtasia™ to capture simultaneously their computer screen, discussions and actions. Data were observed, transcribed and annotated for further analysis of the students' ARwC. Data that were significant to this article were translated from Hebrew to English. Differences of meanings between Hebrew and English connotations of words were discussed extensively, to make sure the original intention of the speaker is clear.

The analysis process focused on the students' ARwC, using interpretative micro-genetic method (Siegler 2006). We examined the entire cohort of data and narrowed it down to reasoning aspects that assemble a narrative of the students' emerging ARwC. Each reasoning aspect is composed of one or more statements of the participants. This process involved many rounds of data analysis sessions and meeting with expert and novice statistics education peers, in which interpretations were suggested, discussed, refined or refuted. This process involved searching forward and

Table 4.1 The Connections 2015 actual learning trajectory

| Activity synopsis | Statistical ideas and concepts | No. of lessons |
|--|--|----------------|
| The first cycle: several contexts (favorite movie, usage of technology, favorite snack etc.) | | |
| Activity 1: Introduction—modelling and data analysis of the class (n = 25 cases) | | 4 |
| Ask a question, plan, collect data from the whole class, model a conjecture about the target population, analyze and organize data, present investigations (hands-on) | The investigative cycle, Uncovering children's initial concepts of reasoning with data, modelling and IIR | |
| Activity 2: Data analysis of the class (n = 25 cases) | | 2 |
| Investigate a data sample from activity 1 (25 cases 2 variables, TinkerPlots™), present investigations | How to ask a statistical question, The investigative cycle, initial reasoning with data, modelling and IIR | |
| Activity 3: The favorite snack activity—modelling and data analysis of the class (n = 25 cases) | | 2 |
| Investigate data sample (25 cases, 3 variables), model a conjecture about the target population (TinkerPlots™), present investigations | Initial reasoning with data, distribution, comparing groups, modelling and IIR | |
| Activity 4: The favorite snack activity—modelling and data analysis of the grade (n = 73 cases) | | 2 |
| Investigate the same question (activity 3) on a larger sample (73 cases, 3 variables, TinkerPlots™), model a conjecture about the target population (hands-on), present investigations | Initial reasoning with data, distribution, comparing groups, modelling and IIR | |
| The second cycle: the amazing race activities | | |
| Activity 5: Data collecting of the entire fifth grade | | 2 |
| Introduction to the activity context, the electronic questionnaire, measurements errors, data collection | How to answer a statistical question, Awareness to the importance of precise data | |
| Activity 6: Modelling and data analysis (n = 10 cases) | | 3 |
| Model a conjecture about the target population (hands-on), investigate data sample (10 cases, 18 variables, TinkerPlots™), refine model, present investigations | Correlations, Noise in the data as a result of measurement errors, The meaning of the median, reasoning with modelling | |

(continued)

Table 4.1 (continued)

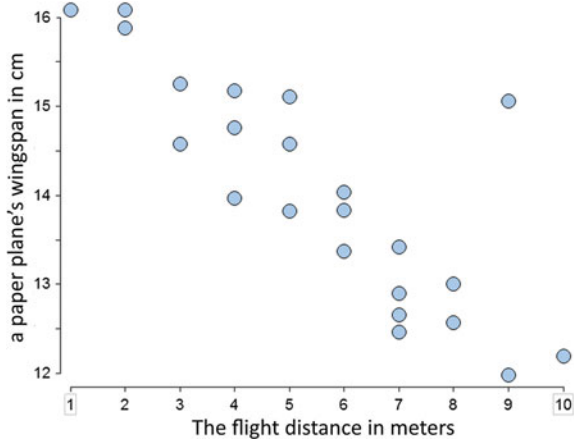
| Activity synopsis | Statistical ideas and concepts | No. of lessons |
|--|---|----------------|
| Activity 7: Modelling and data analysis of the class (n = 24 cases) | | 2 |
| Investigate the same question (activity 6) on a larger sample (24 cases, 18 variables, TinkerPlots™), refine the previous model (hands-on), present investigations | Growing samples, reasoning with data, distribution, comparing groups, modelling and IIR | |
| Activity 8: Modelling and data analysis of the grade (n = 62 cases) | | 2 |
| Investigate the same question (activity 7) on a larger sample (62 cases, 18 variables, TinkerPlots™), refine the previous model (hands-on) | Growing samples, reasoning with data, distribution, comparing groups, modelling and IIR | |
| Activity 9: Modelling of the target population | | 4 |
| Model the conjecture about the target population (TinkerPlots™ Sampler) | Reasoning with data, modelling and IIR | |
| Activity 10: Summarizing | | 5 |
| Prepare a presentation of the investigation and the target population final model, present investigations and models in a student-parents event, write a final report with parents | Reasoning with data, modelling and IIR | |
| Total number of 45-min lessons | | 28 |

backward over the entire data to find acceptable evidences for the researchers' local interpretations and hypotheses (e.g., Ben-Zvi and Arcavi 2001). To strive for "trust-worthiness" (Creswell 2002), inferences about students' reasoning were called only after all data sources (interviews, TinkerPlots™ files and students' notes) provided sufficient evidence, and interpretations from different theoretical perspectives and by a number of researchers were examined (Triangulation, Schoenfeld 2007).

4.5 Results

In this, we present and explain Orr and Guy's emergent ARwC while they made informal inferences and modeled the population. We identify and characterize four reasoning aspects of the students' ARwC in their learning progression. The reasoning aspects varied according to: (a) the analysis unit the students used to examine covariation (for example, a single case, a small group of cases, etc.); (b) the way the students reasoned with signal and noise; and (c) accounted for variability within and between attributes. These aspects represent the key stages of the students' reasoning and helped us to follow the complex process of students' flow of ideas, hesitations,

Fig. 4.1 A relationship between paper planes' wingspan (cm) and their flight distance (m)



mistakes and inventions. The generalizability of this suggested categorization needs to be further studied. Before the description of the four reasoning aspects, we provide the results of the pre-test analysis, and at the end we provide the results of the post-test analysis (the pre- and the post-tests were identical).

4.5.1 Initial Aggregate Reasoning with Covariation (Pre-test)

We analyzed relevant questions from Orr and Guy's pre-tests to reveal the students' initial perceptions of data and variability. In the first question, the students were asked to describe the height distribution of all fifth grade students in Israel, Orr and Guy wrote a single value (145 cm). Later at the same question they were asked to describe the distribution of students' heights in a typical fifth grade class (about 30 students). Orr suggested a range (135–160 cm) while Guy suggested specific height values attached to names of students in his school's fifth grade. In the third question, a scatter plot of a relation between a paper plane's wingspan and its flight distance was presented (Fig. 4.1). The students were asked what the graph described and what can they learn about the relation between paper plane's wingspan and flight distance. Then, they were asked to find where a paper plane of 14 cm wingspan would arrive. Guy's answers were related to only one attribute of the relation. He did not suggest any additional data case. Orr described the relation using the language: "the more... the more". He speculated that a 14 cm wingspan plane would land at the same place as a plane with the exact same wingspan that already appeared in the graph.

Thus, we learn that the students' reasoning with data and variability at the beginning was not aggregative. They held a local view of distributions and although Orr articulated aspects of data aggregation in covariation ("the more... the more"), it seems that his perception of variability was partially interpolating a value deter-

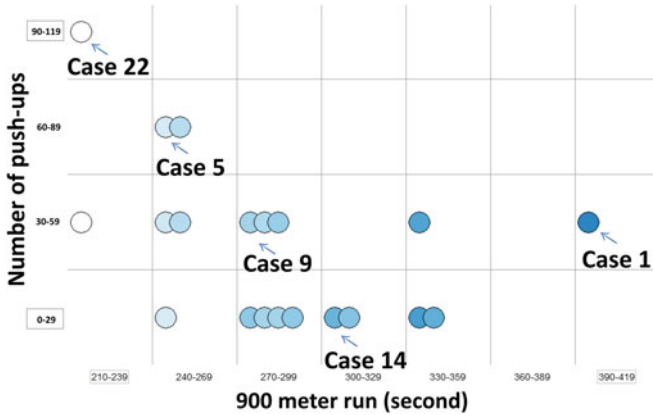


Fig. 4.2 Noticing an initial trend line in the class sample ($n=24$)

ministically. Thus, the students' responses in the pre-test can be considered “single aspects” responses (Moritz 2004).

4.5.2 Aspect 1: A Pointwise-Based Covariation Model

In Activity 7, Orr and Guy extended their previous investigation about the running distribution to study the relations between running and push-ups (a sample of 24). They struggled with formulating an aggregate research question in the first task that reflected their contextual knowledge as well as the essence of the relation. Orr suggested to ask whether push-ups influenced the 900-m run results, while Guy rejected the idea of dependency and suggested asking: “Do other sports relate to running?” The students then drew a scatterplot in TinkerPlots™ and discussed the relations in the data (Fig. 4.2).

- 132 Guy We saw that 93 [push-ups, Case 22] also made a good running [time].
 133 Int. That a person who did many push-ups...
 134 Guy Many push-ups, which is 93, so the [running] result is also good. ...
 143 Int. Do you see, that the higher the push-ups is ...
 144 Orr It gives less... The [running] result is lower.
 148 Guy Here [Case 5], it [push-ups] went down a little, and the [running] result is lower; also here [Case 9] the [running] result is lower, and it [push-ups] went down further more [Case 14] ...
 149 Orr [Continues] and the [running] result is lower. So let's draw a line. [They eventually did not draw this trend line.]
 152 Guy And here [Case 1], It [push-ups] is also quite low, and the lowest [worst running] result.

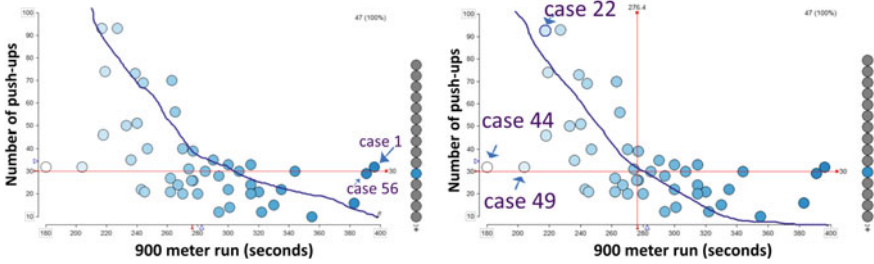


Fig. 4.3 a The first trend line and the outliers. b The second and “reasonable” trend line

Although the students responded to the researcher’s efforts to encourage them to articulate an aggregative expression about the bivariate data [143], they looked mostly locally at the data, starting from an extreme case [Case 22]. They discussed variability by attending to the attributes’ values of each of the five cases (Fig. 4.2) [e.g., 134] and to the value change in relation to the previous case [“it went down a little”, 148]. To provide evidence for covariation, they identified a collection of four points [Cases 22, 5, 9 and 14]. These points made a pattern—a *pointwise-based covariation line*, which they planned to draw but eventually did not [149]. They were possibly inspired by a previous class discussion about trends. Attention to an outlier [152], which did not fit their line [Case 1, worst running result], led them to relate to the noise in the data. We consider their ARwC reasoning at this juncture as a *pointwise-based covariation model*.

4.5.3 Aspect 2: An Area-Based Covariation Model

In Activity 8, following the growing samples method, the students studied a larger sample of the entire fifth grade (47 cases out of 62 cases). They generated a scatter plot, added the means, medians and a horizontal reference line for the median of the push-ups (Fig. 4.3a). Guy drew a descending trend line using the TinkerPlots™ pen and commented that there were cases not represented by this line [Cases 1 and 56 in Fig. 4.3a]. He drew a new “reasonable” [89] trend line in the middle of the data cloud, mostly without passing through cases, but rather between them, and added a vertical reference line of the median running time, which separated the graph into four quadrants (Fig. 4.3b).

The students explained their findings while describing the relation in each of the quadrants.

- 122 Guy The more you approach the height here [at the upper-left quadrant in Fig. 4.3b] ... in the push-ups, the more you progress here, besides maybe these ones [Cases 44 and 49] ... When you are in this area [the upper-left quadrant], it means that he did a lot [a good result] in the running ... If

you are here [Case 22], it means that you do a lot of push-ups and also a good running result.

123 Orr But if you are here [the lower-right quadrant], it means that you ran ...

124 Guy [interrupting] It means that it is quite little push-ups

125 Orr And the result [of the running] is slow.

Orr noticed then that there is another area type in this graph that was created by the intersection of the two median reference lines. Guy drew a circle around it (Fig. 4.4a) and explained that this area is “the center of all these [cases]. The [trend] line that connects between all these [cases] is about the center of the entire [fifth] grade. These are the average [students] of the entire grade” [130]. Later on, the students expanded their view of the covariation and refined their model without any further prompting by the researcher. They speculated about hypothetical data: “If there was one [kid] here [upper-right quadrant], he would not be part of this [attributes relations], since he does many push-ups and a low [running] result” [Guy, 219]. The students summarized their informal inference by saying that “the faster a person runs, the more push-ups he does” [Orr, 134]. When they compared between their class (n=24) and their grade (n=62) samples, they used their new model (Fig. 4.4a) to construct a new representation for their class data (Fig. 4.4b) and were surprised by the similarity between the two samples (Fig. 4.4, b). Although their trend line did not pass through the intersection of median lines, the implementation of their model led the students to discover that the medians’ locations are related to the trend line. They concluded that the medians “show you where the line passes” [Orr, 147].

The students thus reasoned at this juncture with, what we term, an *area-based covariation model*. The combination of the trend line, the four quadrants and the area around the medians’ intersection constructed an *area-based model* for the main features of covariation. This model defined the presence of covariation as a phenomenon in which data cases are located on either upper-left or lower-right quadrants, spread around the signal—the trend line, and sometimes vary a lot (e.g., Cases 44 and 49 in Fig. 4.3b). Outliers were considered as data that appeared on the edges of quadrants. They reasoned with three aspects of the trend line: (a) the location of the trend line; (b) the trend line representativeness of the covariation; and (c) the trend line features (such as, its relation to center measurements). We view the students’ cautious atten-

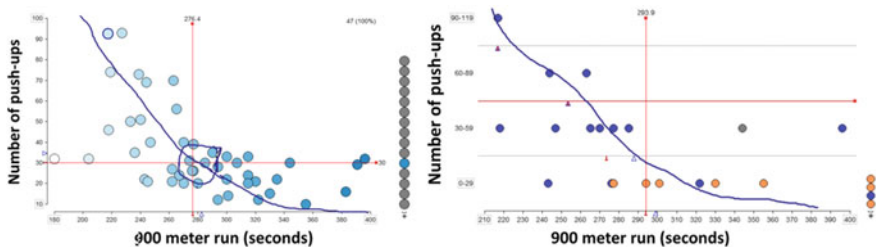


Fig. 4.4 a The relation between push-ups and running in an area-based covariation model (n=62).
 b The relation between push-ups and running colored by gender—the class data sample (n=24)

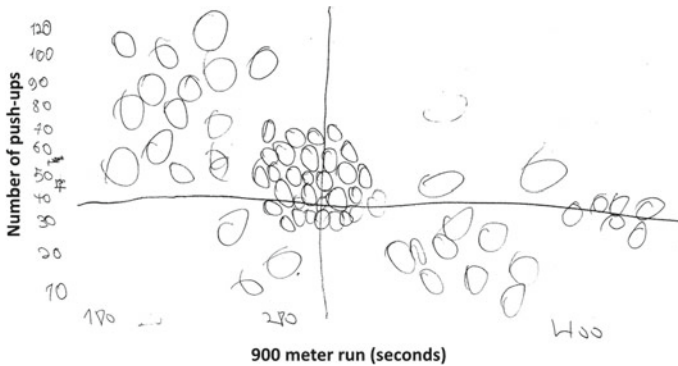


Fig. 4.5 Orr and Guy's conjecture about the target population

tion to these aspects as an attempt to summarize covariation in a way that attended to a large amount of cases and as part of their struggle to view bivariate data aggregately. They reasoned with variability between and within attributes, in relation to a prototype case of a certain area [122]. Using their *area-based covariation model*, they talked about covariation in a rather advanced language: “the more... the more” [122].

4.5.4 Aspect 3: A Cluster-Based Covariation Model

The next task in Activity 8 was to draw a conjecture about the population of all fifth graders in Israel (“the target population”) on a piece of paper. The students followed the researcher’s advice to first describe the push-ups distribution and drew a normal-shaped distribution. This time, they chose not to duplicate the data representation from the TinkerPlots™ real data graph, as they did previously. Rather, they spent time reasoning with aspects of the distribution, considering the data at hand and their beliefs. Later on, they drew the covariation between the attributes and explained their conjectured graph (Fig. 4.5): “[The graph] will be about the same as this one (Fig. 4.4a). There will be many in the middle [the medians intersection]. There will be a many here [the upper-left quadrant] and here [the lower-right quadrant]. About the same amount here and here [upper-left and lower-right quadrants], and here [in the center] also a lot, more than the two of them [upper-left and lower-right quadrants]” [Guy, 323].

The students expressed their conjecture about the target population using, what we term, a *cluster-based model of covariation* (Fig. 4.5). They related to the whole data in a way that expressed covariation between the attributes, by presenting the bivariate data in three clusters with common properties of size and density. They accounted for the variability in the data by noticing the signal and the noise. The signal was the pattern of the correlation, its trend and shape, in terms of the three

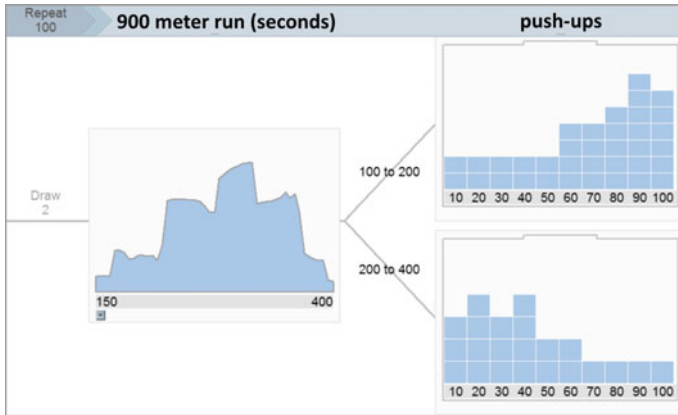


Fig. 4.6 A model of the running–push-ups relations in the population

clusters. They related to outliers and cases at the edges of the quadrants as noise. However, by clustering, they did not relate to the continuity of the aggregate.

4.5.5 Aspect 4: Conditional Distribution Model of Covariation

In Activity 9, the students constructed a model of their target population conjectures (Fig. 4.6) using the TinkerPlots™ Sampler. They first reasoned with the shape and range of the running distribution and defined it as a symmetric “tower with small steps.” They then modeled the dependency between the attributes by separating the running range to two (100–200 and 200–400 s) and explained their choice: “A champion runner will run [900 meters] in a minute and a half, which is about a hundred [seconds]. ... There is no chance that someone [in our sample] will run in a minute and a half” [Orr 26]. They added: “If you ran fast, then you also made a lot of push-ups. If you ran slowly–you made [less push-ups]” [Guy, 37]. The students thus constructed a model of the relations between the attributes while considering variability within and between them, and both data and context.

We term these actions and reasoning as a *conditional distribution view of covariation*, as they described the dependent attribute as two distinct skewed distributions conditioned on the values of the independent attribute. The students’ analysis unit in this aspect was the whole data. Signal was described in relation to the analysis unit and both attributes, while considering continuity in the data. Noise was attended while reasoning with the range, shape, center and tendency of each distribution.

4.5.6 *Articulations of Aggregate Reasoning with Covariation (Post-test)*

We present briefly the results of the post-test, which was identical to the pre-test, to evaluate progress in the students' ARwC. In the first question, both students considered the center, shape and spread of the investigated phenomenon while drawing a height distribution of all fifth grade students in Israel. In the third question, Orr wrote that the graph (Fig. 4.6) described a relationship between planes' wingspans and their flight distance in meters. Both students used an aggregative language to describe this relation as: "the more... the more". When they speculated about a possible flight distance value for a certain plane's wingspan value, Orr interpolated the data case considering variability. Guy considered the center as well, as he explained: "most of the planes are there. Therefore, I thought that [the suggested area] was the average". According to the post-test analysis, the students reasoned with the distribution as an aggregate. Moreover, it seems that there was a progress in their ARwC. When they reasoned with the relation, they described it aggregately and considered aggregate aspects of the relation, such as variability, center and spread.

To sum up, we identified four reasoning aspects that describe the progression of the students' ARwC. In the beginning they described covariation through single cases only. Next, they reasoned separately about areas in the graph, while considering carefully a representative signal within the noise. Their conjecture about the population extended the latter aspect by considering the relations between clusters in the data. Finally, the students related to the whole data, considering continuity and variability in it as well.

4.6 Discussion

This research aims to study the characteristics of the emergent ARwC of two fifth grade students (age 11) who were involved in modelling activities of bivariate data and drawing ISIs in growing samples investigations. We address this goal by carefully analyzing Orr and Guy's emergent processes of ARwC throughout their learning progression. In the following sections, we first describe the students' emergent ARwC and the theoretical implications of our analysis. We then elaborate on the role of the tool and the design approach followed by the research limitations and conclusions. Our main theoretical and pedagogical lessons from this study are:

1. A suggested four-aspect framework of students' emergent ARwC in a learning environment that involves modelling activities and drawing ISIs in growing samples pedagogy.
2. Reasoning with variability and reasoning with modelling play a role in the development of ARwC.

3. The growing samples method, the generation and refinement of models, and the design for purpose and utility are important elements that can support the emergence of ARwC.

4.6.1 *Aggregate Reasoning with Covariation*

In this case study, we identified four different aspects of students' emergent ARwC. These aspects grew from a "single aspects responses" (Moritz 2004), which we identified in the pre-test stage. The four reasoning aspects depict the students' progress from perceiving covariation as a pointwise-based covariation model, an area-based covariation model, a cluster-based covariation model, to conditional distribution model. These reasoning aspects differ by the ways the students attempted to: (1) define an analysis unit to examine covariation; (2) reason with signal and noise; (3) account for variability; and (4) communicate about the correlations between the attributes (discourse about covariation).

The students initially perceived covariation as a *pointwise-based covariation model* (Fig. 4.2). The analysis unit was a single case, starting from extreme values as the most noticeable signal, and following the descending slope of a pointwise line in selecting additional key cases. Cases that only partly met the defined relation were considered as noise. They reasoned with both variability between and within attributes in relation to single cases. Their discourse related to a single case and the way the attributes behaved with regard to this case. When the students analyzed the bigger sample, a new reasoning aspect had emerged: an *area-based covariation model* (e.g., Fig. 4.4a). The analysis units were four quadrants generated by the median reference lines and the area around their intersection. The trend line was considered to be the signal. Cases that were at the edges of the quadrants were considered as noise and outliers. The students reasoned with variability between attributes and discussed covariation in relation to the way the attributes varied within a prototype case of a certain area, considering all cases in the analysis units. When the students modeled their conjecture about the target population, they extended their previous reasoning aspect to a *cluster-based covariation model*. In this aspect, the students described covariation in three main clusters that have common properties (size and density): (1) the center—the "average students" in the population; (2) the upper-left quadrant—the fast runners who do lots of push-ups; and (3) the lower-right quadrant—the slow runners, who hardly do any push-ups. The analysis unit in this perception was the whole data. The signal was the pattern of the correlation, its shape in terms of clusters and the existence of a trend. Cases at the edge of the quadrants were considered as noise, and variability was discussed in relation to the analysis unit, by attending to both attributes. However, they did not consider continuity in the data. The final reasoning aspect we identified in the students' ARwC was a *conditional distribution model*. In this view, the students described the data as a model of two attributes, where they described one attribute by its conditional distribution given the other. The analysis unit in this perception was the data as a whole. Signal was

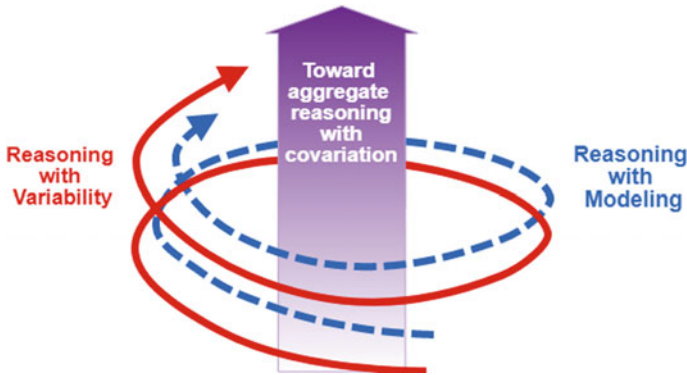


Fig. 4.7 The co-emergence process of ARwC, reasoning with modelling and variability

described in relation to the analysis unit and both attributes, while considering continuity in the data. Noise was attended to while reasoning with the range, shape, center and tendency of each distribution.

It seems, that the emergent process of ARwC involves a progression in the way students view data and covariation. They shift from a local pointwise view of data to an aggregate reasoning with data. Along this journey, they negotiate their understandings of data, variability, center and models to attend to the whole data. This process entails the construction of new understandings of the data at hand, the context of the investigated phenomenon and statistical concepts.

4.6.2 Theoretical Implications

The analysis of this case study distinguished two main processes that seemed to play a central role in the development of ARwC: Reasoning with variability and reasoning with modelling (Fig. 4.7).

4.6.2.1 Reasoning with Variability

We identified in this case study a progression in the students' reasoning in line with the literature (e.g., Garfield and Ben-Zvi 2008; Garfield et al. 2007; Shaughnessy 2007). The students' initial reasoning with variability was first expressed at the pre-test. They hardly attended to variability in data (e.g., represented distribution as single value and did not relate to measurement variability). Later, they related to more informal aspects of variability, such as: identifying that one variable varies more than the other (aspect 2), reasoning with variability with regard to the trend line (aspect 2), reasoning with variability of both variables to reason with the relationship between

the two variables (aspects 3 and 4), reasoning with different representations to view different aspects of variability (aspect 3) and considering measures of variability and center as related while reasoning with data (aspects 3 and 4, and post-test).

We view this process as a key component in the emergence of the students' ARwC. The pointwise-based covariation model (the first reasoning aspect) emerged from concentrating on extreme values and an examination of covariation locally. Such a view toward data restrained the students from reasoning aggregately with data. However, it drew their attention to outliers that did not exactly fit their pointwise-based covariation model. This result is in line with Ben-Zvi and Arcavi (2001) concerning pointwise local view of data and the role of an outlier in developing an aggregate view. On the second reasoning aspect, the students' attention to the variability in data led them to refine the covariation model, i.e., the trend line they drew. We noticed their growing sensitivity to the need to attend to a larger amount of data cases while reasoning with covariation. Their attempt to confront this need was the area-based covariation model. In this model, variability raised the need to justify covariation and characterize each area in relating to all data cases within the certain area. In the cluster-based covariation model (aspect 3), the need to represent covariation led the students to confront variability as they compared clusters in their model and characterized the relations between them. This process extended the analysis unit to the whole data. At the final reasoning aspect, the need to consider the variability of one attributes as depending on the other led the students to extend their ARwC. They considered the whole data, as well as possible interpolations and extrapolations of data, as they constructed the conditional distribution model. They also reasoned with the relation between the center of the distribution and its spread and shape.

4.6.2.2 Reasoning with Modelling

We see the developing process of modelling as another important component in the emergence of ARwC. We assume that each step of a statistical investigation entails a process of emergence, development, refinement or verification of a model (Wild and Pfannkuch 1999). In this case study, some of these models were developed to represent and think or make predictions about the investigated phenomenon (the pointwise line, aspect 1; the trend line, aspect 2; representations of the students' conjectures, aspects 3 and 4). These modelling processes involved attempts to simplify the investigated phenomenon and reason aggregately with data (e.g., Pfannkuch and Wild 2004).

The students' modelling process also entailed the development of the students' epistemological model of the ARwC concept, which we term an *ARwC model*. The emergence of the ARwC model is in line with Ainley et al. (2000) epistemological analysis of trend that includes the sub-elements: correlation, linearity, interpolation and extrapolation. In the context of aggregate reasoning, the students' search for meaning of trend included a search for the relationship between two variables and its representation as a trend. We see this search in the students' request to refine the trend to represent more data (aspect 2), and in their discovery of the relationship

between distributions' centers and the position of the trend (aspect 2). The area-based covariation model extended the ARwC model to a structure of reference and trend lines and a circle. The students used this structure to express and later to examine and evaluate the existence of covariation in data samples of different sizes. We assume that the development of this model facilitated and even promoted the students' perceptions of ARwC to rely on larger amount of data cases as they reason with data. We see the change of the utility of the ARwC model as a transition from a "model of ARwC" to "model for ARwC" (Gravemeijer 1999), as it involved the emergence of a new view of the ARwC concept.

To sum up, this case study implies that reasoning with variability and reasoning with modelling can play an important role in the emergence of ARwC. We see this role as supporting the emergence of ARwC, as well as the growing of understandings of the concepts of variability and modelling (Fig. 4.7). We suggest that this process entails reasoning with the data in hand as well as reasoning with the meaning of statistical ideas (e.g., "model for ARwC"). Further research is needed to study the nature of the roles these aspects play in the emergence of students' emergence of ARwC.

4.6.3 Pedagogical Implications

It seems that the students' learning progression was supported by the design of the learning environment (Ben-Zvi et al. 2018): the growing samples method, the generation and refinement of models and the design for purpose and utility. One of the advantages of the growing samples method is the students' focus on predictions, while viewing these predictions as temporary (see Ben-Zvi et al. 2012). In this case study, the growing samples method elicited the need to summarize data in a way that allows the students to examine their inferences within different size data sets. This requirement brought the need to attend to the signal within the noise. When the students reasoned with different sized samples, they needed to adapt their inferences to a larger data sample. This requirement encouraged them to attend to a larger group of data (e.g., the refinement of the trend line, aspect 2) and later to consider the whole data and possible population (aspects 3 and 4). We assume that the need to model the conjecture about the target population provided a reasonable utility to the data analysis. It also encouraged the students to express their ARwC considering signal, noise and uncertainty (aspect 3) and later dependency and continuity in data in relation to the whole data (aspect 4).

The dynamic TinkerPlotsTM affordance to shift easily between representations helped the students to extend their view and role of the trend line as an aggregate representative of data as a whole (aspects 2 and 4). The TinkerPlotsTM Sampler allowed generating and evaluating different representations, in the search for the one that best expressed the main properties of the investigated concept.

4.6.4 Limitations

This description is far from being a complete description of students' complex emerging processes of ARwC. The two students chosen for this research were considered by their teacher to be both able. This choice was made to enable the collection and analysis of detailed data about their ARwC during the intervention. Even after validating the data interpretation, the idiosyncrasy of the phenomena observed in this research remain questioned. More analysis of students' ARwC should be done within the Connections 2015 learning environment, as well as, further research in other learning environments, to further study students' ARwC.

4.7 Conclusions

This case study presented a new possible learning progression and reasoning aspects of students' ARwC. Students may initially hold local views of covariation. However, when students face covariation in data in such a multi-faceted learning environment, they start considering aspects of reasoning with covariation and develop a sense of the aggregate. Such a reasoning process is involved with handling and confronting variability in data and creating different types of models to analyze data and give meanings to the concept of variability (Fig. 4.7). It seems that this new line of research can advance our ongoing efforts to understand and improve the learning of statistics.

Acknowledgements This research was supported by the University of Haifa and the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation grant 1716/12. We deeply thank the Cool-Connections research group who participated in the Connections project 2015, and in data analysis sessions of this research.

References

- Ainley, J., Nardi, E., & Pratt, D. (2000). The construction of meanings for trend in active graphing. *International Journal of Computers for Mathematical Learning*, 5(2), 85–114.
- Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal*, 32(1), 23–38.
- Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools* (A Ph.D. Thesis). Utrecht, The Netherlands: CD Beta Press.
- Bakker, A., Biehler, R., & Konold, C. (2004). Should young students learn about boxplots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education, IASE 2004 Roundtable on Curricular Issues in Statistics Education, Lund Sweden*. Voorburg, The Netherlands: International Statistics Institute.
- Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distributions. 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., & Hoffmann, M. (2005). Diagrammatic reasoning as the basis for developing concepts: A semiotic analysis of students' learning about statistical distribution. *Educational Studies in Mathematics*, 60, 333–358.
- 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* (pp. 191–205). Voorburg, The Netherlands: International Statistical Institute.
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45(1–3), 35–65.
- 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., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *International handbook of research in statistics education. Springer international handbooks of education* (pp. 473–502). Cham: Springer.
- 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). Berlin: Springer.
- Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events. *Journal for Research in Mathematics Education*, 33(5), 352–378.
- Cobb, P., McClain, K., & Gravemeijer, K. P. E. (2003). Learning about statistical covariation. *Cognition and Instruction*, 21(1), 1–78.
- Confrey, J., & Smith, E. (1994). Exponential functions, rates of change, and the multiplicative unit. In P. Cobb (Ed.), *Learning mathematics* (pp. 31–60). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Creswell, J. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Prentice Hall.
- Dvir, M., & Ben-Zvi, D. (2018). The role of model comparison in young learners' reasoning with statistical models and modeling. *ZDM—International Journal on Mathematics Education*. <https://doi.org/10.1007/s11858-018-0987-4>.
- Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, The Netherlands: Kluwer.
- Friel, S. (2007). The research frontier: Where technology interacts with the teaching and learning of data analysis and statistics. In G. W. Blume & M. K. Heid (Eds.), *Research on technology and the teaching and learning of mathematics* (Vol. 2, pp. 279–331). Greenwich, CT: Information Age.
- 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 teaching practice*. Berlin: Springer.
- Garfield, J., delMas, R. C., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 87–116). New York: Lawrence Erlbaum.
- Gravemeijer, K. (1999). How emergent models may foster the constitution of formal mathematics. *Mathematical Thinking and Learning*, 1(2), 155–177.
- Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic enquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337–364.
- Konold, C. (2002). Teaching concepts rather than conventions. *New England Journal of Mathematics*, 34(2), 69–81.
- 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 (Version 2.0) [Computer software]*. Key Curriculum Press. Online: <http://www.keypress.com/tinkerplots>.

- 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., & English, L. (2017). Introducing children to modeling variability. In D. Ben-Zvi, J. Garfield, & K. Makar (Eds.), *International handbook of research in statistics education*. Springer international handbooks of Education (pp. 229–260). Cham: Springer.
- Lehrer, R., & Schauble, L. (2004). Modelling natural variation through distribution. *American Educational Research Journal*, 41(3), 635–679.
- Lehrer, R., & Schauble, L. (2012). Seeding evolutionary thinking by engaging children in modeling its foundations. *Science Education*, 96(4), 701–724.
- Lesh, R., Carmona, G., & Post, T. (2002). Models and modelling. In D. Mewborn, P. Sztajn, D. White, H. Wiegel, R. Bryant, et al. (Eds.), *Proceedings of the 24th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 89–98). Columbus, OH: ERIC Clearinghouse.
- Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13(1), 152–173.
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82–105.
- Makar, K., & Rubin, A. (2017). Research on inference. In D. Ben-Zvi., K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education*. Springer international handbooks of education (pp. 261–294). Cham: Springer.
- Moore, D. S. (2004). *The basic practice of statistics* (3rd ed.). New York: W.H. Freeman.
- Moritz, J. B. (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.
- 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, Netherlands: Kluwer Academic Publishers.
- Reading, C., & Shaughnessy, C. (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.
- Rubin, A., Hammerman, J. K. L., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In *Proceedings of the Seventh International Conference on Teaching Statistics [CD-ROM], Salvador, Brazil*. International Association for Statistical Education.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32(1), 102–119.
- Schoenfeld, A. H. (2007). Method. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 69–107). Charlotte, NC: Information Age Publishing.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *The second handbook of research on mathematics* (pp. 957–1010). Charlotte: Information Age Publishing Inc.
- Siegler, R. S. (2006). Microgenetic analyses of learning. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (6th ed., Vol. 2, pp. 464–510). Hoboken, NJ: Wiley.
- Watkins, A. E., Scheaffer, R. L., & Cobb, G. W. (2004). *Statistics in action: Understanding a world of data*. Emeryville, CA: Key Curriculum Press.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (with discussion). *International Statistical Review*, 67, 223–265.
- Zieffler, S. A., & Garfield, J. (2009). Modelling the growth of students' covariational reasoning during an introductory statistical course. *Statistics Education Research Journal*, 8(1), 7–31.

Part II
Teaching for Understanding

Chapter 5

Design for Reasoning with Uncertainty



Hana Manor Braham and Dani Ben-Zvi

Abstract The uncertainty involved in drawing conclusions based on a single sample is at the heart of informal statistical inference. Given only the sample evidence, there is always uncertainty regarding the true state of the situation. An “Integrated Modelling Approach” (IMA) was developed and implemented to help students understand the relationship between sample and population in an authentic context. This chapter focuses on the design of one activity in the IMA learning trajectory that aspires to assist students to reason with the uncertainty involved in drawing conclusions from a single sample to a population. It describes design principles and insights arising from the implementation of the activity with two students (age 12, grade 6). Implications for research and practice are also discussed.

Keywords Informal statistical inference · Model and modeling
Sample and population · Statistics education · Uncertainty

5.1 Introduction

Data are everywhere and drawing inferences from data is part of daily life. Every student must therefore have a sense of the potential in drawing reliable statistical inferences from samples, appreciate the purpose of such activity, and deal with the complexities of an uncertain world. However, studies indicate that students can hold contradictory views regarding the relationships between samples and their population (Pfannkuch 2008) and respond in a deterministic way while reasoning about data (Ben-Zvi et al. 2012).

This study is part of the *Connections* Project (2005–2020)—a longitudinal design-based research (Cobb et al. 2003) that studies children’s statistical reasoning in an

H. M. Braham (✉) · D. Ben-Zvi
Faculty of Education, The University of Haifa, Haifa, Israel
e-mail: hana.manor@gmail.com

D. Ben-Zvi
e-mail: dbenzvi@univ.haifa.ac.il

© Springer Nature Switzerland AG 2019
G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_5

inquiry-based and technology-enhanced statistics learning environment for grades 4–9 (Ben-Zvi et al. 2007). The purpose of this chapter is to present the *Integrated Modelling Approach* (IMA)—a pedagogic design approach that intends to help students understand the relationships between samples and populations. More specifically, it describes the design principles of a sixth-grade activity in the IMA learning trajectory and how they contributed to the progression of students’ reasoning with uncertainty.

This chapter begins by describing the instructional design principles of the Connections statistical reasoning learning environment that forms the basis of the IMA learning trajectory. We then address the challenge of facilitating young students’ reasoning with uncertainty while they are involved in making informal statistical inferences (ISI). We describe how the IMA shaped the design of the experimental learning trajectory, provide a detailed description of one activity and illustrate its impact by describing the progression in reasoning with uncertainty of a pair of sixth grade students. This example shows how the students invented methods to face the uncertainty involved in making informal inferences from a sample to a population. Finally, we discuss the challenges in designing activities that foster students’ abilities to envision a process of repeated samples (Shaughnessy 2007; Thompson et al. 2007).

We argue that even relatively young students are able to make sense of complex ideas that form the basis of ISIs, such as, uncertainty and the relationship between data and chance. Furthermore, fostering students’ exploration of two types of uncertainties (contextual and statistical uncertainty) and the connections between them may facilitate students’ understanding of the relationship between a process of repeated samples and a single sample in the inference process.

5.2 Scientific Background

We start this section by describing the design principles of our approach followed by the core statistical ideas of this study—uncertainty in informal statistical inference. Based on these foundations we present the *Integrated Modelling Approach* for supporting the reasoning with sample-population relationships.

5.2.1 Design Principles

Current theories of learning suggest that under certain conditions students who are engaged in carefully designed learning environments may become motivated to construct knowledge from the learning process (Ben-Zvi et al. 2018; Greeno and Engeström 2014). Statistics educators and researchers have recommended the implementation and use of certain statistical learning environments to support the development of students’ statistical reasoning. Garfield and Ben-Zvi (2009) pointed out

several principles of an effective statistical reasoning learning environment (SRLE) to develop students' statistical reasoning. For our study, we adopted four of those principles: Focus on key statistical ideas, use real and motivating data, use inquiry-based activities to develop students' statistical reasoning, and integrate the use of appropriate technological tools.

Focusing on key statistical ideas (such as, distribution, center, variability, uncertainty, and sampling) can stimulate students to encounter them in different contexts and create various representations that illustrate their interrelationships (Garfield and Ben-Zvi 2008). Making connections between existing context knowledge and the results of data analysis can help students develop understanding of key statistical ideas (Wild and Pfannkuch 1999). Using real and motivating data (Edelson and Reiser 2006) through exploratory data analysis (EDA) activities (Pfannkuch and Wild 2004) can help students formulate research questions and conjectures about their explored phenomenon, examine evidence from data in relation to their contextual conjectures, and become critical thinkers in making inferences. Collecting real and authentic data can make the investigation more relevant for students (Herrington and Oliver 2000). Using dynamic visual displays as analytical tools with appropriate technological tools (Garfield et al. 2000) can involve students in the organization, description, interpretation, representation, analysis and creation of inferences of data situations (Ben-Zvi and Arcavi 2001; Ben-Zvi 2006).

5.2.2 *Uncertainty in Informal Statistical Inference*

We first discuss the nature of reasoning with uncertainty in the context of making informal statistical inferences and then consider the challenge of facilitating students' reasoning with uncertainty.

5.2.2.1 Reasoning with Uncertainty

“Statistical inference moves beyond the data in hand to draw conclusions about some wider universe, taking into account that variation is everywhere and that conclusions are uncertain” (Moore 2007, p. xxviii). Given only sample evidence, the statistician is always unsure of any assertion he makes about the true state of the situation. The theory of statistical inference provides ways to assess this uncertainty and calculate the probability of error.

Students, even at a relatively young age, should have a sense of the power and purpose in drawing reliable statistical inferences from samples. Given that statistical inference is challenging for most students (Garfield and Ben-Zvi 2008), Informal Statistical Inference (ISI) and Informal Inferential Reasoning (IIR) became a recent focus of research (Pratt and Ainley 2008; Makar et al. 2011). ISI is a data-based generalization that includes consideration of uncertainty and does not involve formal procedures (Makar and Rubin 2009, 2018). IIR is the reasoning process that

leads to the formulation of ISIs that includes “the cognitive activities involved in informally drawing conclusions or making predictions about ‘some wider universe’ from patterns, representations, statistical measures, and statistical models of random samples, while attending to the strength and limitations of the sampling and the drawn inferences” (Ben-Zvi et al. 2007, p. 2).

Uncertainty is at the heart of formal and informal statistical inference. To understand the uncertainty involved in taking a sample, one needs to envision a process of repeated sampling and its relation to the individual sample (Arnold et al. 2018; Saldanha and McAllister 2014). However, research suggests that students tend to focus on individual samples and their statistical summaries instead of the distribution of sample statistics (Saldanha and Thompson 2002).

5.2.2.2 Facilitating Students’ Reasoning with Uncertainty

Given the importance of IIR, a significant question is how to facilitate young students’ reasoning with uncertainty during sampling and making ISIs. The literature contains examples of two types of settings that have been frequently used to study IIR: (1) scientific inquiry learning environments in which students create surveys and are engaged in real world data inquiries to learn about a wider phenomenon (e.g., Ben-Zvi 2006; Lehrer and Romberg 1996; Makar et al. 2011; Makar and Rubin 2009; Pfannkuch 2006); (2) probability learning environments in which students are engaged in manipulating chance devices such as spinners to learn how probability is used by statisticians in problem solving (e.g., Pratt 2000).

The first setting has considerable potential for students to improve their use of data as evidence to draw conclusions. When students study topics close to their world in an authentic and relevant activity, they can gain important insights into how statistical tools can be used to argue, investigate, and communicate foundational statistical ideas. These settings can also sensitize students to the uncertainty involved in drawing conclusions from samples and the limitations of what can be inferred about the population. However, these settings may lack probabilistic considerations, which contribute to understanding the uncertainty involved in making inferences from samples to populations.

The second setting can encourage and support reasoning with uncertainty. When students manipulate chance devices they can easily build probability models of the expected distribution and observe simulation data generated by the model. They can then compare simulation data with empirical data to draw conclusions. This comparison strategy introduces students to the logic of statistical inference and the role of chance variation. Probability settings, however, may lack aspects of an authentic data exploration and exclude the relevance of the situation.

We suggest that integrating these two settings in making ISIs is important to further support students’ reasoning with uncertainty during sampling. Therefore, we developed an *Integrated Modelling Approach* (IMA) aimed to help students understand the relationships between sample and population. Before presenting the IMA, we first present our conceptual framework for reasoning with sampling.

5.3 An Integrated Modelling Approach for Supporting Sample Population Relationships

5.3.1 Suggested Conceptual Framework for Reasoning with Sampling

We developed an initial conceptual framework (Fig. 5.1) that represents reasoning with sampling during ISIs. According to this framework, reasoning with sampling is an integration of two types of reasoning: (1) reasoning within samples to infer to a population; (2) reasoning between repeated samples.

The first type of reasoning, reasoning *within* sample, is the reasoning involved when exploring real sample data. This includes, for example, looking for signal and noise in data, as well as searching for patterns, trends, and relationships between attributes to learn about real world phenomenon in the population. The second type of reasoning, reasoning *between* samples, is the reasoning involved while drawing repeated collections of samples from the population or from a model of the population. This includes, for example, exploration of sampling variability and examination of the role of sample size on sampling variability. According to this framework, reasoning with sampling creates connections and integration between these two types of reasoning, for example, the relationship between the sampling variability and the likelihood of a single sample statistic.

Our study design was motivated by the hypothesis that integrating between the two types of reasoning with sampling may stimulate students to face both contextual and statistical uncertainty. Contextual uncertainty is the situation in which people are unsure about their context knowledge. The contextual uncertainty stems from a conflict between context knowledge and the sample data at hand. Such a conflict may

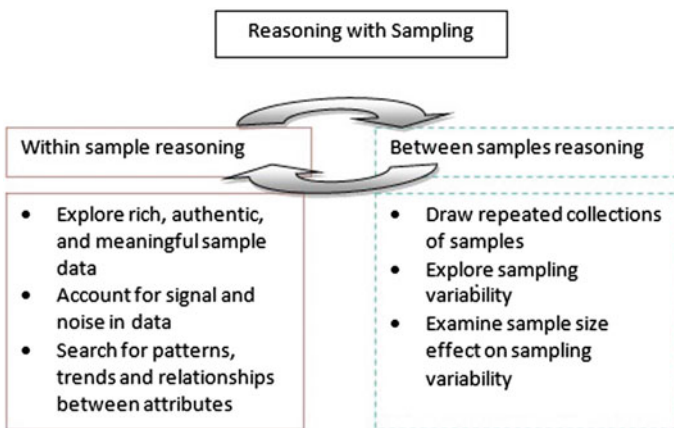


Fig. 5.1 A framework of reasoning with sampling

affect confidence about context knowledge and the ability to infer from a sample at hand. When students infer from a sample to a population their contextual uncertainty may be expressed by probabilistic language (Makar and Rubin 2009) using phrases like: “might be”, “it seems” or “sort of” or by suggesting a subjective confidence level. The subjective confidence level is how certain one feels about inferences in a numeric level (e.g., a number from 1 to 10 or a percent from 0 to 100), which is not calculated but based on subjective estimation. Statistical uncertainty is a situation in which people are unsure about sampling issues such as the behavior of random samples. The statistical uncertainty can be examined and even quantified. For example, the behavior of random samples can be examined by observing sampling variability, and confidence level can be quantified by calculating the probability of getting a statistic as extreme as or more extreme than a specific result, given a specified null hypothesis.

5.3.2 *The Integrated Modelling Approach (IMA)*

Based on these ideas, an *Integrated Modelling Approach* (IMA) was developed by us to guide the design and analysis of a learning trajectory aimed at supporting students' IIR. It is comprised of *data* and *model* worlds to help students learn about the relationship between sample and population. The data world is designed to foster reasoning *within sample*, and the model world is designed to foster reasoning *between samples* (Fig. 5.1).

In the *data world*, students collect a real sample by a random sampling process to study a particular phenomenon in the population. In this world, students choose a research theme, pose questions, select attributes, collect and analyze data, make informal inferences about a population, and express their level of confidence in the data. However, they may not account for probabilistic considerations, such as the chance variability that stems from the random sampling process.

In the *model world*, students build a model (a probability distribution) for an explored (hypothetical) population and generate random samples from this model. They study the model and the random process that produces the outcome from this model. The details vary from sample to sample due to randomness, but the variability is controlled. Given a certain distribution of the population, the likelihood of certain results can be estimated.

In the IMA learning trajectory, students iteratively create connections between the two worlds by working on the same problem context in both worlds and by using TinkerPlots (Konold and Miller 2011). TinkerPlots is dynamic interactive statistics software developed to support young students' statistical reasoning through investigation of data and statistical concepts. The dynamic nature of this software encourages learners to explore data in different repeated representations while testing various hypotheses. TinkerPlots includes a “Sampler” , that allows learners to design and run probability simulations to explore relationships between data and chance, by

means of one technological tool (Konold and Kazak 2008). For a detailed description of the IMA approach see Manor Braham and Ben-Zvi (2015).

5.4 Method

5.4.1 *The Research Question*

This chapter focuses on the design of one activity (The Hidden Model of Social Networks—HMSN) in the IMA learning trajectory that serves as a scaffold for bringing the two worlds closer to the students and fosters students' reasoning with uncertainty. The focus is on the question: *How can reasoning with uncertainty be promoted in a way that is meaningful for young students while they are making ISIs?* More specifically we ask about the role the HMSN Activity played in the development of reasoning with uncertainty in students.

5.4.2 *Methodology*

To address this question, we carried out an illustrative case study of two sixth grade students. We explored their reasoning with uncertainty while making ISIs under the design principles of the activity. Data collection included student responses, gestures (captured using Camtasia), and artifacts (e.g., data representations drawn by them), as well as researcher's observations. All students' verbalizations were carefully transcribed. Interpretive micro-analysis (e.g., Meira 1998), a microgenetic method (Chinn and Sherin 2014), was used to analyze the data. It is a systematic, qualitative, and detailed analysis of the transcripts, which takes into account verbal, gestural, and symbolic actions within the situations in which they occurred. The validation of the data analysis was performed by a small group of statistics education researchers (including the co-authors). The researchers discussed, presented, advanced, or rejected hypotheses, interpretations, and inferences about the reasoning and articulations of the students. The goal of such an analysis was to explore articulations of uncertainty by the students. Initial interpretations grounded in data were reviewed by the researchers and triangulated by a group of expert and novice peers. During these triangulation meetings, hypotheses that were posed by the researchers were advanced or rejected, until a consensus was reached. In order to achieve the necessary "trustworthiness" (Lincoln and Guba 1985), triangulation was achieved only after multiple sources of data validated a specific result (Schoenfeld 2007).

5.4.3 *The Participants*

This study involved a pair of boys (grade 6, aged 12), Shon and Yam, in a private school in northern Israel. The students were selected due to their superior communication skills that provide a window into their statistical reasoning. They participated in a *Connections* unit in fifth grade when they collected and investigated data about their peers using TinkerPlots. Following the growing samples heuristic (Ben-Zvi et al. 2012), they were gradually introduced to samples of increasing size to support their reasoning about ISI and sampling.

5.5 The Hidden Model of Social Networks (Hmsn)

To put the HMSN activity in context, we provide a general description of the entire learning trajectory as well as the rationale and place of the HMSN activity in the learning trajectory.

5.5.1 *The Entire IMA Learning Trajectory*

The learning trajectory¹ encompassed eight activities that initially introduced the two worlds separately. In the data world, the students planned a statistical investigation where they chose a research theme, posed research and survey questions, formulated a conjecture, and decided about the sampling method and sample size (Activity 1). Shon and Yam decided to study the use of technological tools among fourth to ninth grade students in their school. Both Shon and Yam played a lot of computer games, and their research choice arose from their desire to convince the school headmaster to authorize playing computer games at school. Shon and Yam suggested that there were some types of computer games, which they called “wise games,” that can develop thinking and therefore may potentially have a positive influence on students. They decided to explore the relationship between two attributes: whether a student is attentive and whether a favorite type of computer game is “wise.” However, they suggested a biased sampling method of taking two students from each class in grades 4–9 by asking the teachers to choose one attentive child and one non-attentive child. Therefore, we added an activity (Activity 2) to explore the meaning of biased sampling versus random sampling. We also used this activity to expose students to the idea of sampling distribution. The students refined their sampling method, reformulated their conjectures, and implemented a survey in their school (Activity 3). They explored their real data using TinkerPlots (Activity 4). In the model world, they used the Sampler in TinkerPlots to build a hypothetical model for the population distribu-

¹The actual IMA learning trajectory can be viewed at <http://connections.edtech.haifa.ac.il/Research/theimalt>.

tion based on their conjecture. They drew samples from this model, compared them to the model and their real sample data, and explored sampling distributions (Activity 5). To encourage them to examine the connections between the worlds, they were asked “what if” questions on hypothetical real data results while exploring generated random samples.

Since students found it difficult to connect between generated random samples and the real sample, they were given a sixth activity (HMSN Activity 6, which is the focus of the current study). They were asked to study a hidden TinkerPlots model, built by other students, and explore random samples drawn from this model. They then returned to their own investigation and once again explored different sample sizes drawn from their model to compare between them and decide about the minimal sample size needed to draw conclusions about the population. According to their chosen sample size, the students collected more data (Activity 7). Finally, they simultaneously explored data and models in the two worlds by examining the real larger sample data in relation to their conclusions in the model world. They used their estimation of the likelihood to get a specific result given a sample size and a certain distribution of the population, in their conclusions about the population from the real larger sample data (Activity 8).

5.5.2 Rationale of the HMSN Activity

The shift from the data world (Activity 4) to the model world (Activity 5) was challenging for the students. The motivation for the students to move to the model world was that in the model world they would be able to explore two issues: (a) the relationship between random sample and population; (b) the minimal sample size that provides for reliable inferences about the population of interest. While the students explored random model-generated samples they became confused between model-generated samples and real samples. It was challenging for the students to understand what they can gain from exploring the random model-generated samples and how it can help them in investigating real samples. Therefore, we designed a scaffolding activity, the HMSN Activity, to provide a practical purpose for students to study the behavior of many model-generated samples and connect between the repeated sampling and the inferences that are based on a single sample.

5.5.3 The HMSN Activity

A hidden Sampler is a TinkerPlots software option that locks the Sampler to keep students from changing any of its settings and to prevent them from revealing the contents of hidden population devices. In the Hidden Model of Social Networks (HMSN) Activity, students are asked to use the Sampler in TinkerPlots to draw many random samples from a hidden Sampler (Fig. 5.2) to make ISIs. The hidden

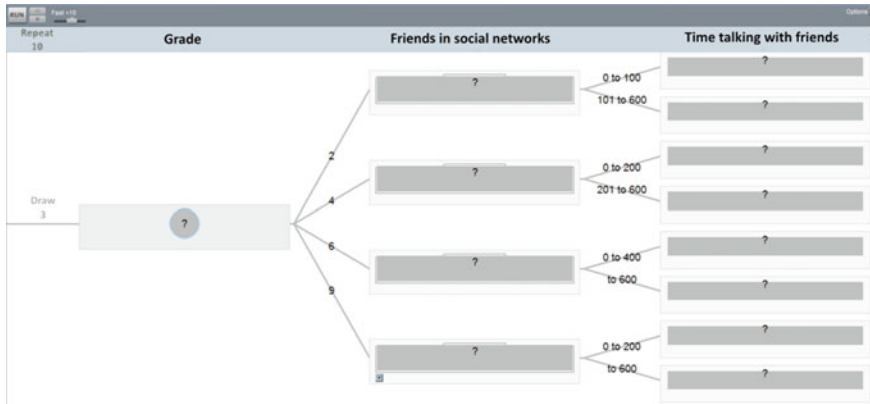


Fig. 5.2 The TinkerPlots Sampler hidden model in the HMSN Activity

Sampler contained three interconnected attributes concerning teenager use of social networks: grade, number of friends in social networks (#FSN), and average time spent on social networks (minutes per day). The students were asked to draw ISIs based on growing sample sizes, beginning their exploration with a small sample size of 10. Each time they wanted to increase the sample size, they had to explain the rationale.

The rationale for including a hidden population was to make the sample-population relationship resemble these relationships in real situations in which the population is unknown. We decided to restrict students to drawing relatively small samples so they would notice the large sampling variability and explore ways to reduce it. We thought this small sample restriction would seem reasonable to the students, since they were aware of the necessity to understand the technical statistical issue of making inferences based on small samples (Ainley et al. 2015). Those students were aware from previous activities of the fact that in real situations one could not collect all data but needed to make inferences on populations from samples. Unlike real life, in the HMSN Activity, the students were able to draw many random samples of a chosen size, and gradually increase the sample size to discover the minimal sample size that can be used for reliable inferences. We hypothesized that following engagement in the HMSN activity, it would be easier for students to enter the model world and the required probabilistic reasoning in the fifth activity.

5.6 Key Features of the HMSN Activity

The main goal of the activity was to develop reasoning with uncertainty of students engaged in sampling during the process of making ISIs. We wanted to motivate and support students in the development of ways to describe, control, and quantify

the uncertainty involved in making ISIs from a single random sample. Guided by the IMA, this activity aimed to support a smooth transition between two types of reasoning: reasoning within samples that occurs in the data world and reasoning between repeated samples that occurs in the model world. In this section we explain the activity design principles and describe the concepts and situations we hypothesize may play a central role in students' reasoning with uncertainty.

5.6.1 Cognitive Conflict Between Data and Context

Most of the theoretical models developed to explain conceptual change (e.g., Strike and Posner 1985; Chi et al. 1994) emphasize the role of cognitive conflict as essential component for conceptual change. Cognitive conflict is generated by dissatisfaction with existing concepts and ideas (Posner et al. 1982). It occurs when a learner cannot use his existing knowledge to solve a problem or explain a phenomenon and is therefore motivated to learn new concepts and ideas (Lee and Kwon 2001). In the learning of statistics, conflicts that take place between former knowledge and current understanding of data analysis can give rise to uncertainty about the explored phenomenon. This can foster and result in new statistical understandings to reduce uncertainty, for example by looking for more data or considering other intervening variables.

Our supposition was that creating conflicts between sample data and context knowledge may motivate students to move from within-sample reasoning (in the data world) to between-samples reasoning (in the model world). To create conflicts we reasoned that an exploration of real and meaningful data was essential. We wanted to ensure that students will easily recognize data that contradicts their experience and be motivated to explore and explain the contradiction.

Shon and Yam were enthusiastic computer users and therefore deeply interested in the theme of this Activity. The research theme they chose in Activities 1-4 was the use and benefit of technological tools among fourth to ninth grade students. The data was also real for the students since the hidden sampler was built by two other students in their class. Those other students built the model while keeping in mind real data that they collected in their school. Therefore, we expected that Shon and Yam would be interested to explore the hidden model data and be equipped with knowledge about its context.

We assumed the students would struggle with sample data that did not make sense in relation to their context knowledge. We hoped that in order to find solutions to those conflicts and handle the uncertainty in data, students would use the TinkerPlots option of generating more samples or consider increasing the sample size. To increase the incidence of conflicts between data and context, the students were asked to begin their explorations with a small sample (size 10). We hoped the students would notice the "noise" (Konold and Kazak 2008) in the data and be motivated to handle the uncertainty by repeated sampling and increased sample size.

5.6.2 *Growing Samples*

Growing samples is an instructional approach mentioned by Konold and Pollatsek (2002), developed by Bakker (2004, 2007) and elaborated by Ben-Zvi (2006). According to the growing samples method students explore small data sets to infer about a wider set of data. They are gradually given more data and asked what can be inferred regarding the bigger sample or the entire population. Therefore, students learn about the limitations of what can be inferred by the teacher's "what-if" questions. This instructional approach was found fruitful in supporting students' reasoning with key statistical concepts such as distribution, variability, and sampling (Ben-Zvi et al. 2012).

In the HSMN Activity, students were exposed to increased sample sizes while expressing considerable uncertainty in small samples due to the limitations of what can be inferred about the hidden Sampler from these small samples. The rationale of using the growing samples heuristic was that it focused the students' attention on inferences (Bakker 2004) and motivated them to develop key statistical ideas and concepts that underlie between-samples reasoning, such as the role of sample size in the confidence level or the connection between sample size and sampling variability.

5.7 Learning Progression of Students

We identified three main thematic and chronological stages in Shon and Yam's expressions of uncertainty: *examine*, *control*, and *quantify uncertainty*. During these stages students gradually refined their way of thinking about uncertainty while learning to integrate the data and model worlds. Due to conflicts they identified, within certain sample results, between data and context knowledge, the students invented these stages (examine, control or quantify) to deal with uncertainty. In this section we describe each one of those stages detailing the conflict and the methods invented by the students to tackle their challenges.

5.7.1 *Stage I: Examine Uncertainty*

5.7.1.1 The First Conflict

Before Shon and Yam drew the first sample (size 10) from the hidden sampler, their initial conjecture was that older students would have more friends in social networks. Observing the sample data, the students were puzzled since the data was in contradiction to their hunch and prior knowledge regarding friends in social networks. For example, they noticed that a fourth grade student had the highest #FSN and that a ninth grade student had no #FSN. Shon commented: "something doesn't make

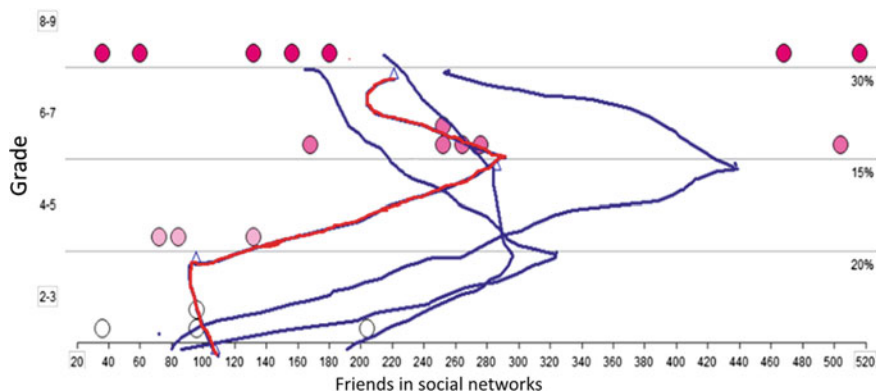


Fig. 5.3 The students' comparison of the MTLs of #FSN over four samples of size 20. A dot represents a case, a blue triangle represents the mean number of friends in a grade, a line connects the four means in each sample, and the red line is the MTL of the fourth sample

sense.” Grappling with this *contextual uncertainty*, the boys added means of #FSN for every grade to find a signal in the data. In their search for a pattern in the data, they used TinkerPlots' drawing tool to connect the means to one another with a “*mean trend line*” (MTL).

5.7.1.2 Students Invent the First Method to Compare Between Samples

To motivate Shon and Yam to consider repeated sampling, we asked them: “What will happen if you drew another sample? Could it help you somehow?” Yam said, in response: “We can take one more [sample],” and Shon excitedly added: “Yes, yes, let's do it [draw from the Sampler] many times.” The students began drawing additional random samples from the Sampler hidden model. To examine the uncertainty caused by the sample variability, Shon and Yam invented a “*Capture MTLs*” method. They plotted the MTL for each sample they drew (Fig. 5.3), compared their position, and noticed the large variability between them. Struggling with this *statistical uncertainty*, Yam reflected: “It [the MTL] is very different each time.” The students consequently asked to increase the sample size from 10 to 20, and Shon stated that, “a sample size of ten is too small.”

5.7.1.3 Reasoning with Uncertainty for Stage I: Examine Uncertainty

During this stage, the students grappled with two types of uncertainty: the contextual uncertainty that stems from a conflict between the data and their prior knowledge as well as statistical uncertainty that stemmed from the large sampling variability they observed and their inability to control random samples. In order to deal with the

contextual uncertainty, they added the MTL to find the signal in the data. In order to deal with the statistical uncertainty, they asked to increase the sample size.

5.7.2 *Stage II: Control Uncertainty*

5.7.2.1 **Second Conflict**

The boys used the same method, “*Capture MTLs*”, to examine the variability between the MTLs of #FSN for samples of size 20. After drawing three random samples from the Sampler, they noticed the smaller variability between the MTLs in comparison to samples of size 10. During these explorations of samples (size 20), they referred to the similarities and differences in location, shape and “peak” (maximal mean) between the MTLs. They noticed a trend in the #FSN (The number of friends increased from grade 2 to 7 and decreased from grade 7 to 9). They explained that the reason for the decrease from grade 7 to 9 is that ninth graders have usually more homework and exams and therefore have less time to communicate with friends on social networks. However, they still expressed their statistical uncertainty and wanted to further increase the sample size.

A fourth MTL surprised them (the red line in Fig. 5.3) and destabilized their relative confidence regarding the MTL’s trend (e.g., unlike the previous samples, the number of friends decreased from grade 2 to 5 in the fourth sample). Yam said, “It [this fourth sample] is very bad.” Instead of drawing more samples, they asked to increase the sample size once again. At this point, the researcher tried to motivate the students to draw more samples by asking: “Do you feel more confident in your conclusions about certain grades?” In response, the students decided to draw many samples of size 20 and examine the variability between the means of #FSN within the grades.

5.7.2.2 **Students Invent a Second Method to Compare Between Samples**

The students developed a new graphical method, “*Capture Means*”, to capture the variability between the means in order to examine whether they could control the uncertainty in the repeated sampling process. According to their “*Capture Means*” method, when the mean result of a particular grade could be captured inside a drawn circle, they concluded that the variability within that grade was small. They drew several samples of size 20, and Yam noticed that in grade 6, “It [the mean variability] is relatively stable because it [the mean] is usually in the area of the circle. That’s why I say that they [the means over several samples, Fig. 5.4] are relatively stable.” However, the students noticed that the mean results from the three other grades could not be captured inside a drawn circle. Therefore, they expressed higher statistical uncertainty regarding the sample size and the resulting conclusions. Due

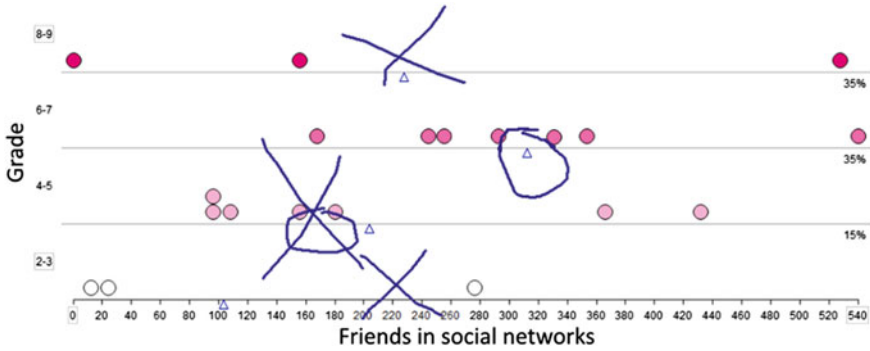


Fig. 5.4 “Stable” and “constantly varying” mean signals over several samples of size 20

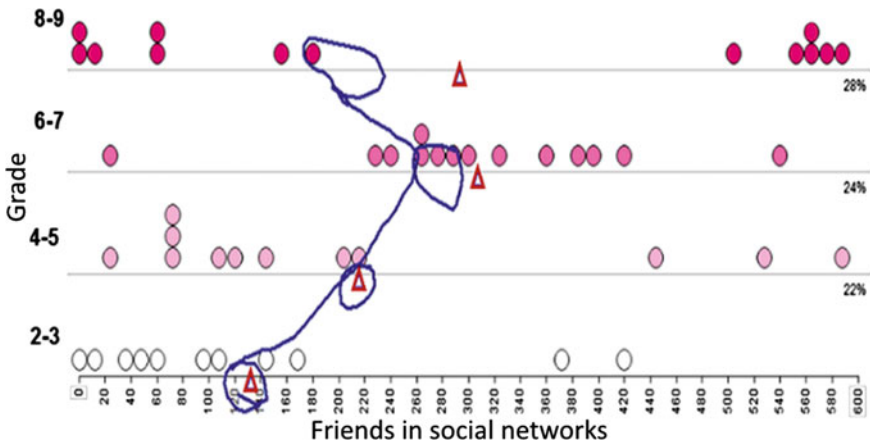


Fig. 5.5 The hypothetical MTL of #FSN over several samples size 50

to the outcome of only one circle (i.e., stable mean) over many samples, they asked to increase the sample size to 50.

Shon and Yam applied their “*Capture Means*” method on larger samples of size 50, drew a circle for each grade capturing the means of that grade over many samples (Fig. 5.5). They noticed that, “grade 9 [means] stay in this area [the top blue circle in Fig. 5.5]. It [grade 4 means] really jump around this spot [Yam drew a circle around grade 4 means].” Encouraged by these results, the boys expressed a higher confidence level and were satisfied with the sample size. Shon said: “In my opinion, [sample of] 50 will be enough.” Their confidence about the MTL’s stability increased, and they connected the four circles (Fig. 5.5) saying they were “absolutely certain.”

5.7.2.3 Reasoning with Uncertainty for Stage II: Control Uncertainty

In the beginning of the second stage, comparing samples (size 20) and observing smaller variability between MTLs, decreased the students' statistical uncertainty. A surprising sample showing an MTL that was incongruous with their former knowledge increased their contextual uncertainty causing them to feel unsure about their context knowledge. To handle the contextual uncertainty, they decided to invent the "*Capture Means*" method and examine whether they could control the means of #FSN in certain grades. Examining several samples of size 20 with the "*Capture Means*" increased their statistical uncertainty regarding conclusions that could be drawn from random samples of size 20. To deal with the statistical uncertainty, they increased the sample size to 50 and with the "*Capture Means*" method decreased their statistical uncertainty. To examine their conclusions in relation to their former knowledge, they drew a new MTL. The similarity between the MTL and their hypothesis decreased their contextual uncertainty.

5.7.3 Stage III: Quantify Uncertainty

5.7.3.1 Third Conflict

During the next meeting, the students' confidence encouraged them to refine their hypothetical MTL for samples of size 50. They drew a few random samples but were surprised that several of them showed a significantly different trend than the hypothetical trend. They therefore decided to differentiate between two main trends: "type 0" trend (the MTL of #FSN is increasing between grade 2 to 7 and is decreasing between grade 7 to 9) and "type 1" trend (the MTL of #FSN is decreasing between grade 2 to 5, increasing between grade 5 to 7 and increasing between grade 7 to 9). They complained: "We can't draw an inference because it is different all the time." They tried to deal with the growing uncertainty about the trend by drawing bigger random samples of size 65 and noticed that there were more samples of "type 0" than "type 1" trend.

5.7.3.2 Students Invent a Third Method to Compare Between Samples

To *quantify* their uncertainty about the trend, the boys invent a third method to compare between samples. They calculated the difference between the numbers of samples within each trend, and called this difference a "breakpoint." For example, if the first and second samples showed "type 0" trend and the third sample showed "type 1" trend, they said that the breakpoint is one (2-1). They decided that when this breakpoint equals a certain number, determined in advance, it would point at the more likely trend. Setting the breakpoint to three, the boys strengthened their previous assumption and chose "type 0" trend over "type 1" estimating their subjective

confidence level to be 80%. They explained their high confidence level in “type 0” trend by referring to the difference between the number of samples with “type 0” trend to those with “type 1” trend. They even found a way to increase their confidence level in “type 0” trend by setting the breakpoint to five.

Yam: Because we had three times more [cases of “type 0” than “type 1”]. There are still times it’s like this [“type 1”], but most of the time it’s like this [“type 0”].

Shon: We will wait until it [the breakpoint] will be more than five. Here again... we’ll wait until it will arrive at five... If there’s one more time [a sample with “type 0” trend], then I believe 90% [that the trend is of “type 0”].

At this point they generalized the meaning of the breakpoint to be an estimate of their confidence level; a bigger breakpoint results in a higher confidence level.

5.7.3.3 Reasoning with Uncertainty for Stage III: Quantify Uncertainty

In the third stage the students felt unsure about their context knowledge because they observed some MTLs that were incongruous with their former knowledge, a fact that increased their contextual uncertainty. Furthermore, they felt unsure about the ability to infer from random samples of size 50, a fact that increased their statistical uncertainty. To deal with the uncertainties, the students increased the sample size to 65 and quantified the sampling variability by calculating the difference between the number of samples in each trend. However, to express their level of confidence in their inference, they didn’t make calculations but used a subjective confidence level of 80%.

5.8 Discussion

The main question of this chapter is: How can reasoning with uncertainty be promoted in a way that is meaningful for young students while they are making ISIs?

In this chapter we presented the IMA and the design principles of one activity in the IMA learning trajectory. Our analysis illustrated how the students’ reasoning with uncertainty was refined during their engagement in this activity. In the following section we discuss the research conditions and its limitations followed by the pedagogical and theoretical implications of our analysis regarding the relationship between: a) one sample and repeated samples, and b) data and chance. We also discuss our main design challenges to highlight the characteristics of the activity that cultivated the progress of the students learning.

5.8.1 *Limitations*

The purpose of describing the design of a learning trajectory and an analysis such as the one presented in this chapter is that researchers and teachers could learn from it and adjust the activities to their circumstances. Therefore, it is important to provide the conditions and limitations of this research. This chapter is based on our analysis and experience with only a small number of students who had superior communication and reasoning skills. Therefore, our findings are only a proof of principle. More research is needed to determine how this activity in particular and the IMA in general can be performed with less intervention in a classroom setting. We are currently conducting another study with sixth grade students in a classroom setting to test the idiosyncrasy and the generality of the case presented in this chapter.

Researchers and teachers will also need to consider that our students were involved in EDA activities during the previous year. Throughout that year we exposed them to ideas of sample size and inferences that can be drawn from a sample. We think that in the IMA learning trajectory, students' experience with an exploratory approach to data is essential for entering the model world and dealing with the complex idea of uncertainty (Pfannkuch et al. 2012). Reasoning with uncertainty in the context of informal statistical inference is an ongoing discourse aimed to convince others regarding inferences that can be made and the level of confidence in making those inferences. The fact that our students were used to an environment of open and critical discourse from the previous year prepared them to deepen their reasoning with uncertainty and inferences this year.

5.8.2 *Implications*

5.8.2.1 **Pedagogical Implications**

Our case study demonstrates how reasoning with uncertainty developed through students' iterations between the data and the model worlds. In our analyzed data, the students' expressions of contextual and statistical uncertainties shaped their movement between the worlds. The *contextual uncertainty*, which occurred in the data world, stemmed from the conflict between the boys' context knowledge and the data in relation to a specific sample. For example, when the boys explored a sample size of 10, Shon doubted that a fourth grade student had the biggest #FSN and thought that "it is strange." Such a conflict increased the boys' uncertainty about the ability to infer from a sample. The *statistical uncertainty*, which occurred in the model world, stemmed from sampling variability. Disconcerted by small sample sizes and restricted by the activity design, the boys invented graphical methods to examine the variability between means and MTLs over many samples. These situations increased the boys' uncertainty about the ability to infer from a single sample of a certain size.

In order to understand the uncertainty involved in taking a sample, one needs to envision a process of repeated sampling and its relation to the individual sample (Saldanha and McAllister 2014). The relationship between the individual sample and repeated samples may emerge during the construction of the relationship between contextual and statistical uncertainties. Furthermore, articulations of statistical uncertainty may emerge from the need to face and explain the contextual uncertainty and evolve with repeated sampling. For example, the need to elucidate conflicts between data and prior knowledge and the ability of the tool (TinkerPlots) to draw repeated samples assisted students in examining whether the conflicts happened due to chance and impelled them to face statistical uncertainty.

Distinguishing between the two types of uncertainties may be important from a pedagogical point of view. As we depict in this study, facing a contextual uncertainty may motivate students to examine statistical uncertainty by drawing repeated samples and observing sampling variability. Therefore, we argue that designing activities that promote conflicts between data and context knowledge and encouraging students to consider repeated sampling may be fruitful in understanding the relationship between sampling variability and confidence in a single sample.

5.8.2.2 Theoretical Implications

The findings of this study are consistent with the argument that students must be able to integrate between data and chance in order to understand informal statistical inference (e.g., Konold and Kazak 2008; Pfannkuch et al. 2018). This is due to the fact that making ISIs involves connecting probability-based notions of uncertainty and inferences that are drawn from data (Makar and Rubin 2009, 2018). Although researchers agree that EDA is an appropriate method for exploring statistics, a criticism of the EDA pedagogical approach towards informal statistical inference is its data-centric perspective (Prodromou and Pratt 2006) that does not foster students' appreciation of the power of their inferences as does the model-based perspective (Horvath and Lehrer 1998; Pfannkuch et al. 2018). This study responds to the challenge of reconnecting data and chance bi-directionally with an *Integrated Modelling Approach* that adds elements of a model-based perspective to the EDA approach. We suggest that engaging students with iterations between the data and model worlds in the IMA, as presented in the HMSN Activity Section, may help them integrate ideas of data and chance.

Figure 5.6 summarizes the students' iterations between the data and model worlds and between data and chance. Pronounced conflicts between data and context knowledge that were expressed by contextual uncertainty in the data world (the left column in Fig. 5.6) played an important role in the boys' motivation to examine chance, as well as invent and refine their methods of examining, controlling, and quantifying the statistical uncertainty in the model world (the right column in Fig. 5.6). During the first stage, exploring sample data that contradicted their previous knowledge in the data world played an important role in the boys' motivation to move to the model world, draw repeated samples, and invent the "Capture MTLs" method to examine

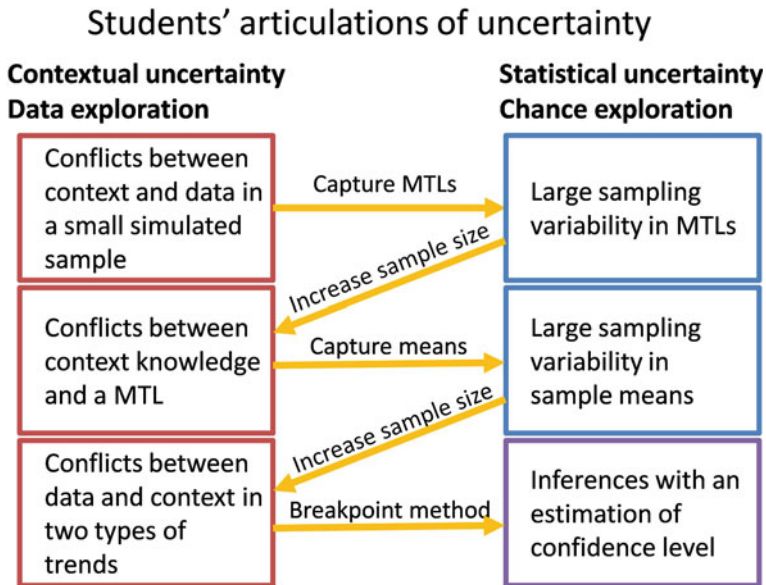


Fig. 5.6 The students' reasoning with uncertainty through iterations between the worlds

the statistical uncertainty in sampling variability. By examining sampling variability, they actually explored whether the conflicts they observed between context and data were due to chance. The large sampling variability compelled them to increase sample size. During the second stage, a surprising sample showing MTL that made no sense in the data world forced the boys to invent the “*Capture Means*” method to control the statistical uncertainty. The quantification of the statistical uncertainty in the third stage resulted from their contextual uncertainty regarding the hypothetical MTL.

Shuttling between the worlds, the students were able to make meaningful connections between inferences they can draw about a phenomenon from samples of a certain size and the idea of repeated samples and sampling variability. Our case study depicts that, in carefully designed activities that cultivate the idea of repeated sampling (Shaughnessy 2007; Thompson et al. 2007), even relatively young students can be exposed to and make some sense of complex ideas behind ISI such as the relationship between sample size, sampling variability, and confidence level in a sample of a certain size.

5.8.2.3 Design Implications

One challenge in cultivating reasoning with uncertainty in the context of ISI is how to motivate students to deal with statistical uncertainty. In other words, how to create situations in which students see utility (Ainley and Pratt 2010) in drawing many

samples to examine uncertainty. Utility of an idea is an understanding what it is useful for and what power it offers in addressing problems with respect to a ‘project’ in which the student is currently engaged. Moving to the model world and envisioning a process of repeated sampling and its relation to the individual sample (Saldanha and McAllister 2014) is not easy or natural for the students (Saldanha and Thompson 2002). One reason for that may be that the idea of repeated sampling is too theoretical for students, and they usually don’t find utility in the action (Ainley et al. 2006). However, in the HMSN Activity, students found utility in drawing many samples and used it to increase their informal confidence level in the inference that could be made from a single sample. Informal confidence level is an estimation of how certain one feels about informal inferences. It is uttered by a numeric level that is not calculated but based on a relative number of repeating samples that indicate a particular result.

We hypothesize that the combination of engaging students with real and meaningful data that motivates them to deal with contextual uncertainty and the possibility in TinkerPlots of drawing many random samples of different sizes, assisted students in comparing between samples.

Although the HMSN Activity included an artificial task and took the students away from their focus project, they were aware of the statistical idea of sampling and the need to examine the power it had on estimating the level of confidence in samples in their ongoing project. We think so since the students were engaged before the HMSN Activity in inquiry-based activities based on real sample data in both the data and the model worlds. During those activities they were dealing with the questions: “Can one trust random samples?” and “What is a sufficient sample size on which one can make reliable inferences on the population?” So in this context, the artificial task in the HMSN Activity was connected to their ongoing project. Furthermore, after the HMSN Activity, the students returned to work on their real data and used what they learned about sampling and uncertainty to find the minimal sample size on which they could make reliable inferences about the explored population. We believe that in such a learning environment, if students find utility in drawing many samples as a way to face uncertainty, there is a greater chance that they will understand this concept and also utilize it in other contexts.

A second challenge was in motivating students to invent methods to compare between samples. We didn’t want to prescribe a comparison solution prior to their experience. Furthermore, we thought that by inventing methods to explore and compare between samples, students would have the opportunity to struggle with the rationale of examining many samples and their relationship to a single sample. We suggest that since these students were used to an exploratory approach to data (EDA), it seemed natural for them to look for and invent different methods to compare between samples. In the previous year and in the first four activities of the IMA learning trajectory, the students looked for different methods to analyze data in order to convince their peers about their inferences. The current activity with its use of TinkerPlots enabled students not only to find methods to analyze sample data but also to draw many samples and find innovative ways to compare between them.

Inventing methods to compare between samples has conceptual consequences. First, observing many samples and deciding how to compare between them can support the concept of aggregate view (Konold et al. 2015; Aridor and Ben-Zvi 2018). For example, while students invented and examined their *MTL method*, they realized the importance of the location and “peaks” of the line, in addition to its shape. Second, while students are engaged in inventing methods to compare between samples, they can learn different ways to notice and describe sampling variability and its relation to sample size.

However, inventing methods to compare between samples invited also meta-conceptual questions such as: How can we compare between samples? What does it mean to compare between samples? What is a good method for comparison? What information is missing in our method? For example, when the boys began to compare the MTLs they realized that there was a similarity between the shapes of the MTLs, but there were differences in the MTLs location and “peaks”. Therefore, they looked for other ways to compare between samples and invented the “*Capture Means*” method that helped them focus on the variability of the #FSN means locations within the grades, over many samples.

Although on a small scale, this study sheds light on new ways to combine data and chance, in order to support students’ informal inferential reasoning. Helping students make connections between data and chance using the IMA pedagogy will inevitably bring with it new challenges regarding learning to make ISIs and smoothing the transitions between the data and model worlds. However, these difficulties can be embraced as essential steps in the development of the reasoning of students who are engaged in a modern society in which drawing inferences from data becomes part of everyday life.

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. (2010). It’s not what you know, it’s recognising the power of what you know: Assessing understanding of utility. In C. Reading (Ed.), *Proceedings of the 8th International Conference on the Teaching of Statistics (ICOTS)*. Ljubljana, Slovenia: International Statistical Institute and International Association for Statistical Education. Retrieved from <http://www.Stat.Auckland.Ac.Nz/IASE/Publications>.
- 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. (2018). Statistical modeling to promote students’ aggregate reasoning with sample and sampling. *ZDM—International Journal on Mathematics Education*. <https://doi.org/10.1007/s11858-018-0994-5>.
- Arnold, P., Confrey, J., Jones, R. S., Lee, H. S., & Pfannkuch, M. (2018). Statistics learning trajectories. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 295–326). Cham: Springer.
- Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: CD-β Press, Center for Science and Mathematics Education.

- Bakker, A. (2007). Diagrammatic reasoning and hypostatic abstraction in statistics education. *Semiotica*, 2007(164), 9–29.
- 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 (CD-ROM)*, Salvador, Bahia, Brazil, July 2006. Voorburg, The Netherlands: International Statistical Institute and International Association for Statistical Education.
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45, 35–65.
- 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., 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 Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL5)*, University of Warwick, UK, August 2007.
- Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 473–502). Cham: Springer.
- Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, 4, 27–44.
- Chinn, C. A., & Sherin, B. L. (2014). Microgenetic methods. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 171–190). New York: Cambridge University Press.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.
- Edelson, D. C., & Reiser, B. J. (2006). Making authentic practices accessible to learners: Design challenges and strategies. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences*. New York: Cambridge University Press.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: 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., Chance, B., & Snell, J. L. (2000). Technology in college statistics courses. In D. Holton (Ed.), *The teaching and learning of mathematics at university level: An ICMI study* (pp. 357–370). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Greeno, J. G., & Engeström, Y. (2014). Learning in activity. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 128–150). New York: Cambridge University Press.
- Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development*, 48(3), 23–48.
- 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.
- 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). <http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art1/>.
- Konold, C., & Miller, C. (2011). *TinkerPlots™ 2.0*. Amherst, MA: University of Massachusetts.
- 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.
- Lee, G., & Kwon, J. (2001). *What do we know about students' cognitive conflict in science classroom: A theoretical model of cognitive conflict process*. Paper presented at Annual Meeting of the Association for the Education of Teachers in Science, Costa Mesa, CA, January 18–21, 2001.

- Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction*, 14(1), 69–108.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- 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.
- Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261–294). Cham: Springer.
- 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, Minnesota: Catalyst Press.
- Meira, L. (1998). Making sense of instructional devices: The emergence of transparency in mathematical activity. *Journal for Research in Mathematics Education*, 29(2), 121–142.
- Moore, D. (2007). *The basic practice of statistics* (4th ed.). New York: W. H. Freeman.
- Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics (ICOTS) [CD-ROM]*, Salvador, Bahia, Brazil.
- Pfannkuch, M. (2008, July). Building sampling concepts for statistical inference: A case study. In *The 11th International Congress on Mathematics Education*. Monterrey, Mexico.
- Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM—International Journal on Mathematics Education*. <https://doi.org/10.1007/s11858-018-0989-2>.
- 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). Netherlands: Springer.
- Pfannkuch, M., Wild, C., & Parsonage, R. (2012). A conceptual pathway to confidence intervals. *ZDM—The International Journal on Mathematics Education*, 44(7), 899–911.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
- Pratt, D. (2000). Making sense of the total of two dice. *Journal for Research in Mathematics Education*, 31(5), 602–625.
- Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 3–4.
- 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.
- Saldanha, L., & McAllister, M. (2014). Using re-sampling and sampling variability in an applied context as a basis for making statistical inference with confidence. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS-9)*. Voorburg, The Netherlands: International Statistical Institute and International Association for Statistical Education.
- Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51(3), 257–270.
- Schoenfeld, A. H. (2007). Method. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 69–107). Charlotte, NC: Information Age Publishing.
- 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 Publishing.
- Strike, K. A., & Posner, G. J. (1985). A conceptual change view of learning and understanding. In L. West & L. Pines (Eds.), *Cognitive structure and conceptual change* (pp. 259–266). Orlando, FL: Academic Press.

Thompson, P. W., Liu, Y., & Saldanha, L. A. (2007). Intricacies of statistical inference and teachers' understandings of them. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 207–231). New York: Lawrence Erlbaum Associates.

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

Chapter 6

Building Concept Images of Fundamental Ideas in Statistics: The Role of Technology



Gail Burrill

Abstract Having a coherent mental structure for a concept is necessary for students to make sense of and use the concept in appropriate and meaningful ways. Dynamically linked documents based on TI© Nspire technology can provide students with opportunities to build such mental structures by taking meaningful statistical actions, identifying the consequences, and reflecting on those consequences, with appropriate instructional guidance. The collection of carefully sequenced documents is based on research about student misconceptions and challenges in learning statistics. Initial analysis of data from preservice elementary teachers in an introductory statistics course highlights their progress in using the documents to cope with variability in a variety of contextual situations.

Keywords Concept image · Deviation · Distribution
Interactive dynamic visualization · Mean · Variability

6.1 Introduction

Educators have suggested that visual images provide an important tool for learning (e.g. Breen 1997). Dreyfus (1991) argued that the “status of visualization in mathematics education should and can be upgraded from that of a helpful learning aid to that of a fully recognized tool for learning and proof” (vol. I: p. 33). Presmeg (1994) suggested that visualizing mathematical concepts is a means to develop understanding. Interactive dynamic technology can be an important factor in helping students build these images. This view is supported by a number of studies that suggest strategic use of technological tools can help students transfer mental images of concepts to visual interactive representations that lead to a better and more robust understanding of the concept (e.g. Artigue 2002; Guin and Trouche 1999). In particular, technology plays a central role in teaching and learning statistics, perhaps a greater role

G. Burrill (✉)

Program in Mathematics Education, Michigan State University, East Lansing, MI, USA
e-mail: burrill@msu.edu

© Springer Nature Switzerland AG 2019

G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_6

123

than for many other disciplines (Chance et al. 2007). A variety of researchers have investigated the role of technology in the learning of statistics (c.f., Ben-Zvi 2000; Burrill 2014; Watson and Fitzallen 2016). In many classrooms, however, the use of technology can too easily focus only on organizing data, graphing and calculating. The perspective throughout this chapter is that technology, particularly interactive dynamic technology, can and should be used for more than “doing the work”, a view supported by the American Statistical Association’s Guidelines for Assessment and Instruction in Statistics Education (GAISE) that stresses the use of technology for developing conceptual understanding as well as carrying out analyses (Franklin et al. 2007).

6.2 The Potential of Interactive Dynamic Technology

Content specific learning technologies provide many opportunities for developing understanding of statistical concepts. Interactive dynamic technology allows students to link multiple representations—visual, symbolic, numeric and verbal—and to connect these representations to support understanding (Sacristan et al. 2010; Biehler et al. 2013; Burrill 2014). For example, a regression line can be dynamically linked to a visualization of the residual squares and the numerical sum of the squared residuals. Such interactive linking, where one object is manipulated and all related representations are instantly updated, supports investigations into varying assumptions and asking “what if” questions that can lead to making and testing conjectures and result in a better understanding of the concepts involved (Ben-Zvi 2000). Computer simulation activities enable students to experience variability by comparing random samples, generating simulated distributions of sample statistics, and observing the effect of sample size on sampling distributions (delMas et al. 1999; Hodgson 1996). The ability to display multiple screens simultaneously allows students to contrast different graphs of the same data or notice how changing a data point affects a distribution. Spreadsheet features provide opportunities for managing large sets of data, enabling students to investigate subsets of the data for similarities and differences, for example, sorting a data set according to gender to compare curfews or spending money. Many misconceptions held by students about statistical concepts can be confronted using technology in a “predict and-test” strategy, establishing a cognitive dissonance that can help students change their thinking about a concept (e.g., Posner et al. 1982). Students can predict what they think they will observe (e.g., expected shape of a distribution) and then use the technology to obtain immediate feedback on their thinking.

From a more fundamental perspective, however, interactive dynamic technology has the potential to help students create robust *concept images* of key statistical ideas, a necessary step in being able to fluently and effectively reason with and apply those ideas (Oehrtman 2008). A concept image can be described as the total cognitive structure including the mental pictures and processes associated with a concept built up in students’ minds through different experiences associated with the

ideas (Tall and Vinner 1981). Without a coherent mental structure, students are left to construct an understanding based on ill formed and often misguided connections and images (Oehrtman 2008). The work of understanding subsequent topics is then built on isolated understandings specific to each topic (e.g., center as separate from spread, distribution as a set of individual outcomes, randomness as accidental or unusual). This makes it difficult for students to see and work with the relationship among the images needed for deep understanding, for example, to understand the distinction among the distribution of a population, the distribution of a sample from that population, and the distribution of a statistic computed from samples from the population.

A *concept definition* can be thought of as the words used to specify the concept, which are typically related in some way to a student's personal experiences with the concept. As students engage in new experiences related to the concept, a student's concept image changes and evolves into a personal concept definition. For example, a student's first image of mean might be "add and divide"—an image of the specific rule for calculating the mean of a set of data. If the concept image of mean remains at this level, students will struggle when they are asked to interpret a mean in context or approximate a mean from the graph of a distribution. The educational goal should be to provide students with experiences that will help them move to a more formal understanding of the concept, supported by the development of rich interconnected concept images/definitions, that is accepted by the community at large (Tall and Vinner 1981).

Piaget argued that an individual's conceptual structure is based on the actions or the coordination of actions on physical or mental objects made by the individual (Piaget 1970, 1985). Given this stance, instruction beginning with formal definitions would seem to be contrary to the direction in which abstraction occurs. Oehrtman (2008) suggests three important features of instructional activities compatible with Piaget's theory of abstraction. First, the underlying structure that is the target for student learning should be reflected in the actions they do. Because these actions come before conceptual understanding, they should be stated in terms accessible to students rather than formal definitions, enabling students' eventual concept images to build from conceptual structures that make sense to them because of their previous actions. Second, students' actions should be repeated and organized with provisions for feedback and ways to respond to this feedback. And third, students should use these actions in structurally similar problems in a variety of contexts to develop a robust abstraction of the concept.

This chapter describes a sequence of applet-like documents, *Building Concepts: Statistics and Probability* (2016) (BCSP) developed according to an "action/consequence principle" aligned with Oehrtman's features (2008). The materials were created to exploit the affordances of an interactive dynamic environment in developing robust conceptual structures for key statistical concepts, designed for teaching introductory statistical concepts.

6.3 Building Concepts

6.3.1 An Action/Consequence Principle

To make sense of ideas, it is necessary to have appropriate conceptual structures, and it is impossible to communicate about concepts without any representations (Bakker and van Eerde 2014). In his semiotic theory on the use of diagrams as ways to represent relationships, Peirce describes *diagrammatic reasoning* as that which involves constructing a diagram, experimenting with it, and reflecting upon the results. He emphasizes that in the experimenting state, “thinking in general terms is not enough. It is necessary that something should be DONE. In geometry, subsidiary lines are drawn. In algebra, permissible transformations are made. Thereupon the faculty of observation is called into play” (CP 4.233—CP refers to Peirce’s collected papers, volume 4, section 233). Because the learner has done something with the diagram or representation, he is forced to consider the consequences of the action from a different perspective than that originally in his mind (Peirce 1932, 1.324).

In *Building Concepts*, these three steps are embodied in an “action/consequence” principle, where the learner can “deliberately take a mathematical action, observing the consequences, and reflecting on the mathematical implications of the consequences” (Mathematics Education of Teachers II 2012, p. 34). In statistics, the actions might involve grouping data points in a certain way, changing bin widths in histograms, moving data points, generating random samples from a population, changing the sample size, or moving a line. The consequences might be different visual representations of the data, changes in numerical summaries, noting what remains constant and what changes with the action, or a shift in patterns. By reflecting on the changes they see in response to statistically meaningful actions, students are engaged in actively processing, applying, and discussing information in a variety of ways (National Research Council 1999; Michael and Modell 2003) and can begin to formulate their own concept images and conceptual structures of key statistical ideas.

From another perspective, the theories of Mezirow (1997), Kolb’s learning cycle model (1984), and the work of Zull (2002) on brain theory all suggest that people learn through the mechanism of participating in an immersive mathematics experience, reflecting on these experiences, and attempting similar strategies on their own. Mezirow introduced the notion of transformative learning as a change process that transforms frames of reference for the learner. Key elements in this process are an “activating event” (Cranton 2002) that contributes to a readiness to change (Taylor 2007). This is followed by critical reflection where the learner works through his understanding in light of the new experiences, considering the sources and underlying premises (Cranton 2002). The third element of this process is reflective discourse or dialogue in an environment that is accepting of diverse perspectives (Mezirow 2000). The final step is acting on the new perspective, central for the transformation to occur (Baumgartner 2001). These four elements elaborate on Kolb’s early model of experiential learning (1984) as a cycle containing four parts: concrete experi-

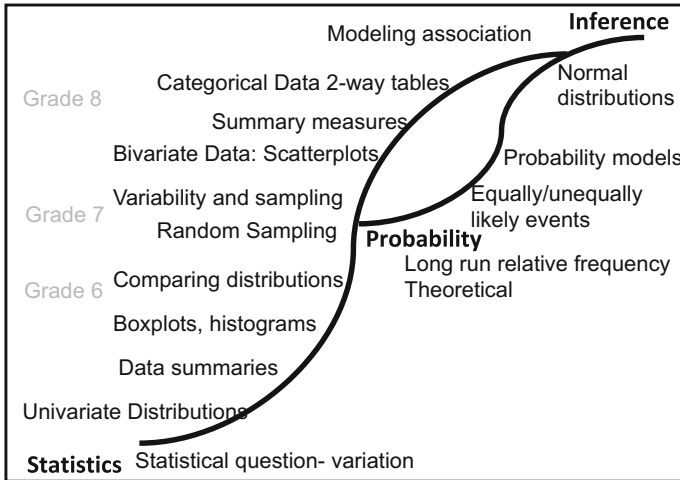


Fig. 6.1 Statistics and probability—a coherent progression

ence, reflective observation, abstract conceptualization, and active experimentation; experimentation leads once again to concrete experience. This cycle, informed by Oehrtman’s key features, is embodied in the action/consequence principle underlying the activities in *Building Concepts*.

6.3.2 Content Framework

The content in the *BCSP* activities is based on the Common Core State Standards (CCSS) progressions documents (2011), narratives describing the learning progression of a topic based on the research on cognitive development and on the logical structure of mathematics/statistics. Taken as a whole, the activities and corresponding dynamic files cover the key concepts typically in introductory school statistics (Fig. 6.1). Static pictures or examples contained in the progression document are made interactive in the activities. In addition, the activities have been designed in light of the research related to student learning, challenges and misconceptions.

6.3.3 The Activities

The core of the activities are applet-like, dynamic interactive files, not intended to be used for “doing” statistical procedures but rather to provide a mental structure for reasoning about statistical concepts that can support the transition to procedural fluency. When students have a solid conceptual foundation, they can engage

in statistical thinking, are less susceptible to common errors, less prone to forgetting and are able to see connections and build relationships among ideas (NRC 1999).

6.3.3.1 Framing of Tasks

The tasks in each activity focus on using the interactive documents to create experiences that can contribute to the development of a particular statistical concept. They were constructed following the advice of Black and Wiliam (1998) with respect to formative assessment: “Tasks have to be justified in terms of the learning aims that they serve, and they can work well only if opportunities for pupils to communicate their evolving understanding are built into the planning (p. 143).” Thompson (2002) argued that the goal of a task is to have students participating in conversations that foster reflection on some mathematical “thing”. Thus, the majority of tasks in the activities create opportunities to discuss particular statistical objects or ideas that need to be understood and to ensure that specific conceptual issues and misconceptions will arise for students as they engage in discussions.

6.3.3.2 Misconceptions

The tasks in the activities have been designed in light of the research related to student learning, challenges and misconceptions (e.g., Zehavi and Mann 2003). For example, a common misconception in statistics relates to boxplots: the longer one of the four sections in the plot, the more data in that section (Bakker et al. 2005). To build a mental image of the connection between the data and a boxplot, in the interactive file a dot plot “morphs” into the boxplot, and students can compare the number of data values in each section of the boxplot (Fig. 6.2). Moving points in the dot plot immediately displays the effect on the corresponding boxplot (Fig. 6.3), reinforcing the fact that medians and quartiles are summary measures based on counting.

The activity *Equally Likely Events* was designed specifically to address the misconception that every outcome has a 50% chance of occurring (Fischbein et al. 1991). In this activity, students generate a distribution of the eleven possible sums of the faces when two dice are tossed and compare the distribution to a distribution of the outcomes of spinning a spinner divided into eleven equal regions. The visualization of the distributions as the number of repetitions is increased makes explicit how a random sample reflects the characteristics of the population. In *Comparing Distributions*, students explicitly contrast histograms and bar graphs to confront the confusion they often have distinguishing between the two representations. They consider the limitations of bar graphs in understanding the story of the typical income, education and life expectancy in various regions of the world and compare what is lost or gained when the data are represented in boxplots, histograms, or dot plots. Students create histograms with a large amount of variability and with little variability to challenge

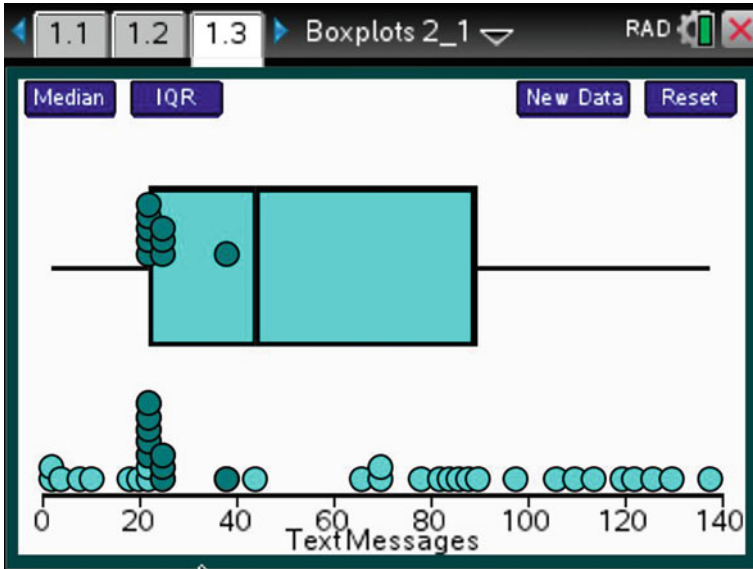


Fig. 6.2 Connecting boxplots to data

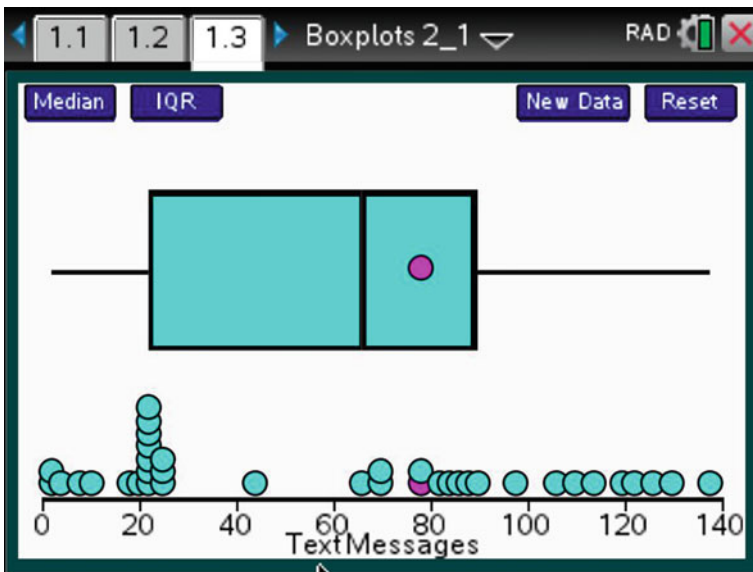


Fig. 6.3 Moving a data point

the misconception that variability is defined by the range or by a peak rather than the spread around the mean (delMas and Liu 2005; Matthews and Clark 2003).

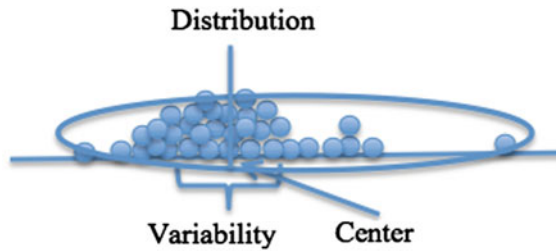
6.3.3.3 Posing Questions

In addition to making sure that the tasks surface misconceptions and develop understanding of “tough to teach/tough to learn” concepts, the questions for each of the activities were created using the general guidelines below:

1. Activate prerequisite knowledge before it is used; e.g., “Remember the importance of thinking about shape, center and spread when talking about distributions of data. Describe the distribution on page 1.3.” (*Introduction to Histograms*)
2. Point out things to notice so students focus on what is important to observe; e.g., “Select Sample. Describe the difference between the points on the number line at the top left and the point on the number line at the right.” (*Sample Means*)
3. Ask for justifications and explanations; e.g., “Make a conjecture about which data set will have the largest mean. Explain why you think your conjecture might be correct. Use the file to check your thinking.” (*Mean as Balance Point*)
4. Make connections to earlier tasks or to an immediately previous action taken by the student (questions should not come out of the blue); e.g., “Return to your answers for question 2 and see if you want to change them now that you have looked at the values when they are ordered.” (*Median and Interquartile Range*)
5. Include both positive and negative examples in developing understanding of definitions, theorems and rules; e.g., “Which of the following are true? Give an example from the Ti-Nspired file to support your reasoning. (a) The smallest and largest values of any distribution are outliers. (b) Not all distributions have outliers. (c) An outlier will be more than one boxplot width plus half of the width of the boxplot to the left and right of the box. d) The segments on each side of the box always extend $1\frac{1}{2}$ IQRs beyond the LQ and the UQ.” (*Outliers*)
6. Have students consider the advantages/disadvantages of each approach when it is possible to carry out a task using multiple strategies; e.g., “Which, if any, of the three estimation methods—educated guess, judgment sample, or random sample—do you think is more likely to give a sample that is most representative of the population? Why?” (*Random Samples*)
7. Be explicit about possible misconceptions: e.g., “Work with a partner to create two reasonable distributions for the number of pairs of shoes owned by the students in a class, either by moving or adding points, to get (1) a distribution with little variability in the number of pairs of shoes owned by most of the class, and (2) a distribution where there is a lot of variability in the number of pairs of shoes owned by the class. Choose a bin width that seems best for your distribution. Describe your distribution (shape, center and spread). Explain why you think one of your distributions has very little variability and the other has a lot of variability.” (*Introduction to Histograms*)

The next section provides several examples of using the dynamic interactive files to develop concept images related to core statistical ideas. These include distributions, measures of center and spread, and random behavior including sampling variability.

Fig. 6.4 Image of distribution, center and variability



6.4 Developing Concepts

6.4.1 Distributions

A statistical distribution might be defined as “an arrangement of values of a variable showing their observed or theoretical frequency of occurrence” (The Free Dictionary). Wild (2006) suggests, however, that “... the notion of “distribution” is, at its most basic, intuitive level, “the pattern of variation in a variable,” (p. 11) and further notes that focusing on what a distribution is will not be as productive as focusing on helping students build a mental image of how data can be distributed. According to Wild, because distributions are such a fundamental component of statistical reasoning the goal should be on how you can reason with distributions and not on how do you reason about distributions.

A well-documented problem observed by statistics educators is that students tend to perceive data as a series of individual cases and not as a whole that has characteristics and properties not observable in any of the individual cases (i.e., Bakker and Gravemeijer 2004; Ben-Zvi and Arcavi 2001; Hancock et al. 1992). They suggest that students need to develop a conceptual structure in which data sets are thought of as aggregates where the concept image of how data can be distributed includes features related to shape, center and variability around the center (Fig. 6.4).

In the first *BCSP* activity, *Introduction to Data*, students investigate numerical lists and dot plots of the maximum recorded speeds and life spans of different animal types (Fig. 6.5) with the goal of building a mental image of a distribution of the data. Students begin by identifying individual animals or data points (How fast is a tiger?). They are then asked to talk about the distribution as a whole, connecting words such as “clumps”, “bumps”, “piles”, “dots are spread out” (Bakker and Gravemeijer 2004; Cobb et al. 2003) to shapes, eventually building images of distributions that can be described using language accepted in the statistical community: mound shaped, symmetric, skewed, uniform. Transitions from language such as “all bunched at one end” to “skewed” or “the dots are spread out” to “the spread is large” are important steps in the formation of concepts (Peirce 1998). The interactive files allow students to notice how changing a data point affects a distribution and to experiment with removing data points in a distribution to see the effect on the shape (How will the

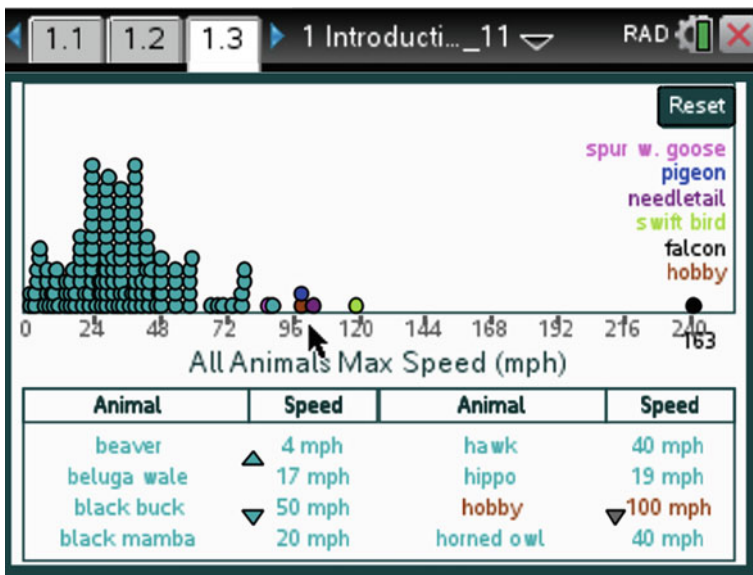
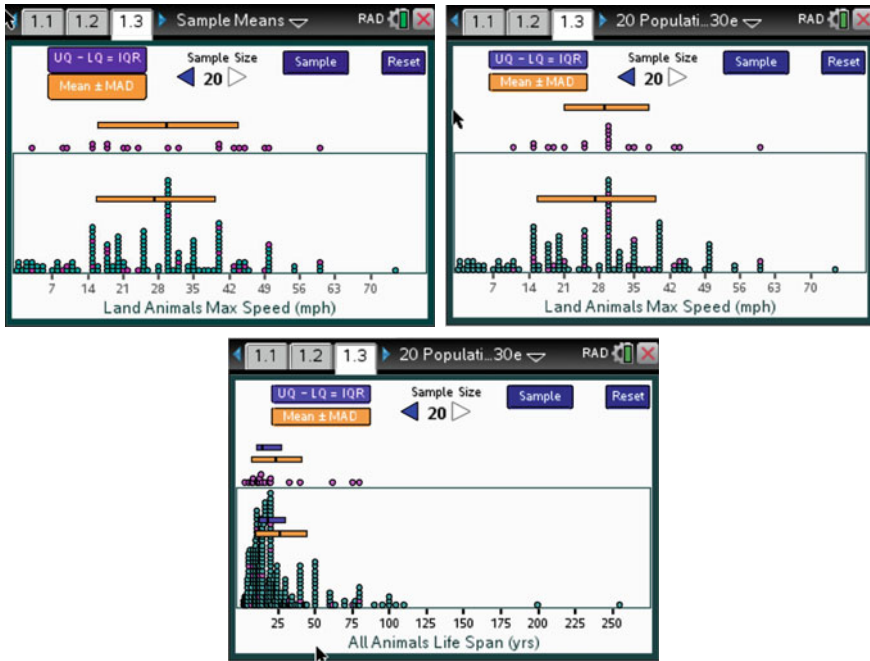


Fig. 6.5 Maximum speeds of types of animals

shape change if the maximum speed of the Peregrine falcon is removed from the distribution? If the speeds for all of the birds are removed?)

Building from the conceptual structures students have formed in this initial work and mindful of the principle that students should use these actions in structurally similar problems in a variety of contexts, the concept of distribution is revisited in other activities, such as those which develop the connection between measures of center and spread and the shape of a distribution. The technology allows students to cycle through a variety of data sets, providing opportunities to recognize distributions when the mean may not represent the largest cluster of data points, and the median may be a more useful measure of center.

The concept of distribution is revisited again in the context of sampling. The conceptual structures students have developed for reasoning with distributions are extended to consider distributions of a sample from a population, where, for example, they examine the distributions of maximum recorded speeds for a sample of animal types, the plot on the top in Figs. 6.6, 6.7, and 6.8, and distinguish this from the distribution of the maximum recorded speeds for all of the animal types, the plot on the bottom in Figs. 6.6, 6.7, and 6.8. Repeatedly taking samples provides contexts in which the distribution of sample maximum recorded speeds reflects the population but varies from sample to sample as do the summary measures (mean \pm mean absolute deviation or median and interquartile range) associated with the random samples of the maximum speeds, reflected by the horizontal bars in Figs. 6.6, 6.7, and 6.8.



Figs. 6.6–6.8 Populations, samples and sampling variability

6.4.2 Mean and Standard Deviation

Students’ concept images of measures of center and spread seem to be fragile. Misconceptions or superficial understanding of measures of center have been well documented (Friel 1998; Groth and Bergner 2006; Mokros and Russell 1995; Watson and Moritz 2000). Students often can perform the computations but cannot apply or interpret the concepts in different situations and have correspondingly ill formed notions of variability. In the past many texts introduced the mean and median as measures of center in a single lesson, and several lessons later or in another chapter, if at all, introduced measures of variability. Treating center and spread together supports the creation of a mental structure of the notion that measures of spread are connected to “spread around what”—some value indicating a measure of center (see Fig. 6.4); deviations are measures of distance from the mean, and the interquartile range (IQR) is a measure of the distance between the first and third quartiles and thus around the median. Experiences with these different interpretations of center and variability, can help students build a mental structure mindful of the need to take both measures into account when reasoning about variation in a variety of situations (Shaughnessy et al. 1999) and can help them recognize that either measure alone tells an incomplete story about the context.



Fig. 6.9 Mean as leveling

In *Building Concepts*, median and interquartile range are introduced in one activity followed by activities related to mean and mean absolute deviation (which is introduced as a precursor to standard deviation). The literature suggests that typically students have problems interpreting the mean and applying it appropriately (e.g., Garfield and Ben-Zvi 2005). To counter this, the activities explicitly develop the concept of mean as “fair share” in two ways. The activities endeavor to build mental images of (1) fair share as “leveling off” where students drag dog food bags from the dogs who have the most bags to dogs with fewer bags (Fig. 6.9) until all of the dogs have the same number of bags; and (2) fair share as pooling, where all of the contributors (the dogs) put their bags of dog food into a group (Fig. 6.10), and the entire group is then divided equally among the total number of contributors (dogs) (Fig. 6.11). Both approaches contribute to developing images needed for complete understanding of mean; the first develops an understanding of how to interpret the mean as a measure of center, and the later leads directly to the procedural algorithm typically used to compute a mean.

Recognizing the difficulty students have shifting their images of bar graphs as ways to describe distributions of data to graphs involving quantitative data displayed on a number line, one file focuses on connecting numerical (16 total bags) and pictorial representations to a dot plot, where students observe how the dot plot changes as the pictorial representations are moved (Fig. 6.12). The fact that all of the dots are in a vertical line at four when each dog has four bags of dog food lays the groundwork for considering the mean as a balance point.



Fig. 6.10 Pooling



Fig. 6.11 Dividing up the pool

Mean As Balance Point is set in the context of soccer tournaments, where the task is to distribute a given total number of goals in a tournament to achieve a mean number

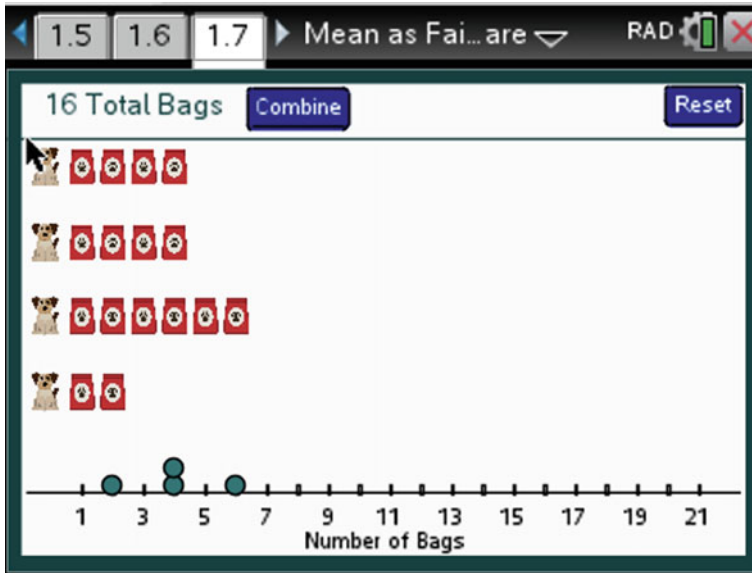


Fig. 6.12 Connecting representations

of six goals for the nine teams involved (Kader and Mamer 2008). An important part of the reflection step in the action/consequence principle is for students to describe what they see in the diagrams (Figs. 6.13 and 6.14) then, with support, learn to abstract from the picture the notion of deviation, where deviation in itself can have characteristics (Pierce 1998). The goal of this activity is to give students experience in describing deviations, resulting in the development of an image for the concept of deviation as an object itself and to eventually link deviation to the concept of variability. Students move dots representing the number of goals scored by a soccer team to “balance” the dots on the number line, given that the mean number of goals for all of the teams has to be six, and certain constraints must be satisfied (e.g., no teams scored six goals, two teams scored two goals, one team scored three goals, and one team scored nine goals). They can notice how changing a data point affects the distribution of goals and explore how the “deviations” from the mean are related to whether the segment containing the distribution of goals is balanced.

Students identify a measure to rank different tournaments (distributions) in terms of “most evenly matched teams” with the assumption that, in a tournament with perfectly matched teams, every team scores the same number of goals (Fig. 6.15). This leads to the mean absolute deviation as a measure of spread around the mean and the notion of mean as balance (Fig. 6.16). The development “uncouples” the words “standard” and “deviation”, giving students the opportunity to build an image of the word deviation in a simple context before they think about standard deviations.

Associating shapes with measures of center and variability can help students develop an understanding of what these measures mean graphically and numerically

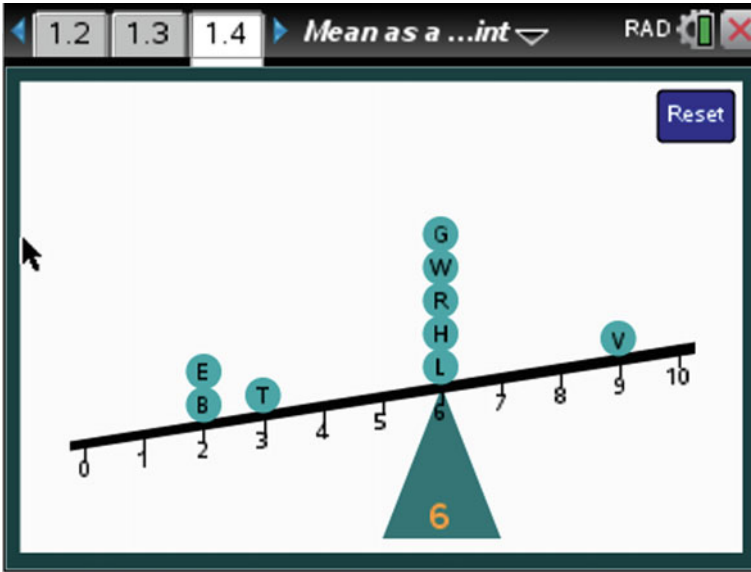


Fig. 6.13 Goals in a tournament

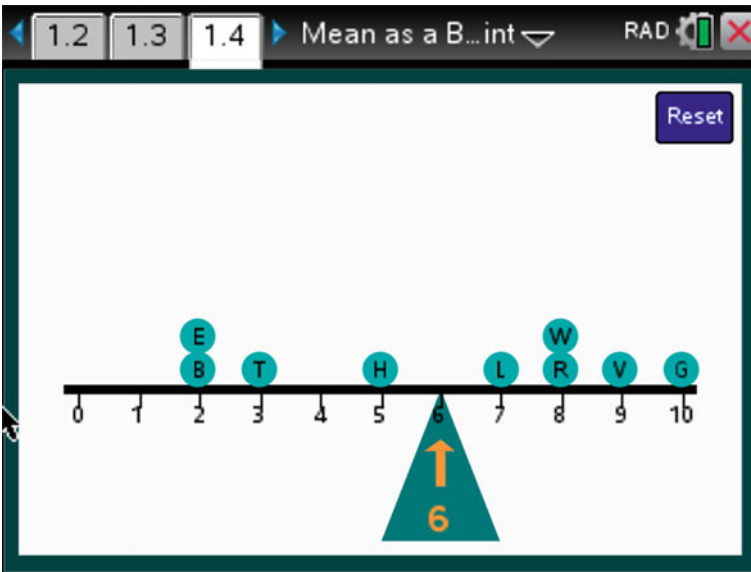


Fig. 6.14 Constraints satisfied

(Garfield and Ben-Zvi 2005). Connecting the image of deviation to a mental image of variation around the mean, students use the technology to make conjectures about

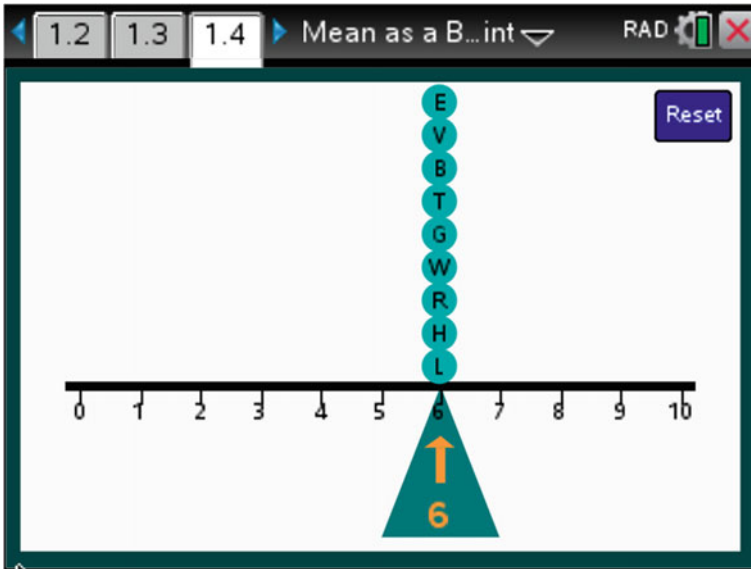


Fig. 6.15 Evenly matched



Fig. 6.16 Ranking soccer tournaments in terms of the most “evenly balanced” teams

the measures of center and spread for randomly generated distributions of scores and can instantly check their conjectures (Fig. 6.17). The technology supports students in continuing to build their mental images by making visible the connections among numerical, visual and algebraic representations as they interpret data in a table and relate the data and summary measures to a graph (Figs. 6.18 and 6.19).

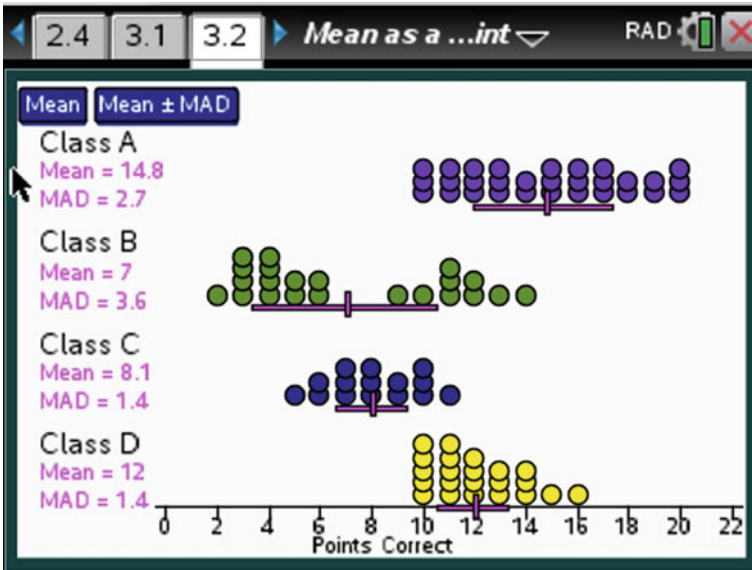


Fig. 6.17 Checking conjectures

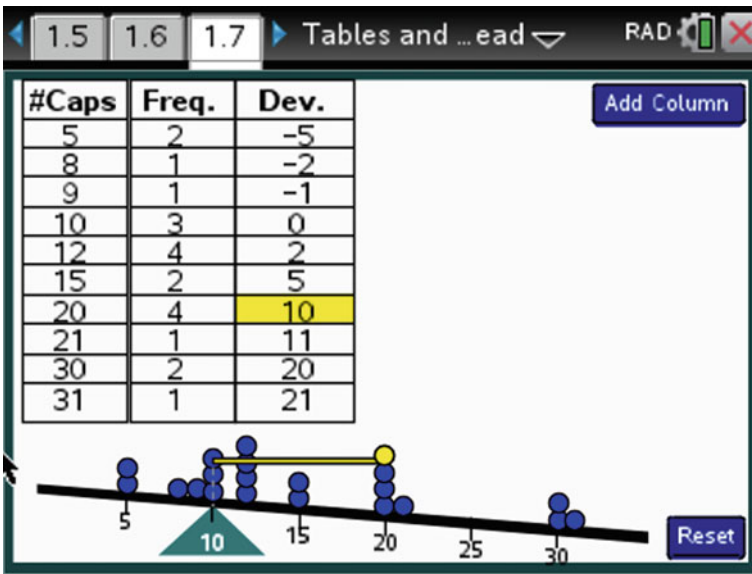


Fig. 6.18 Deviations

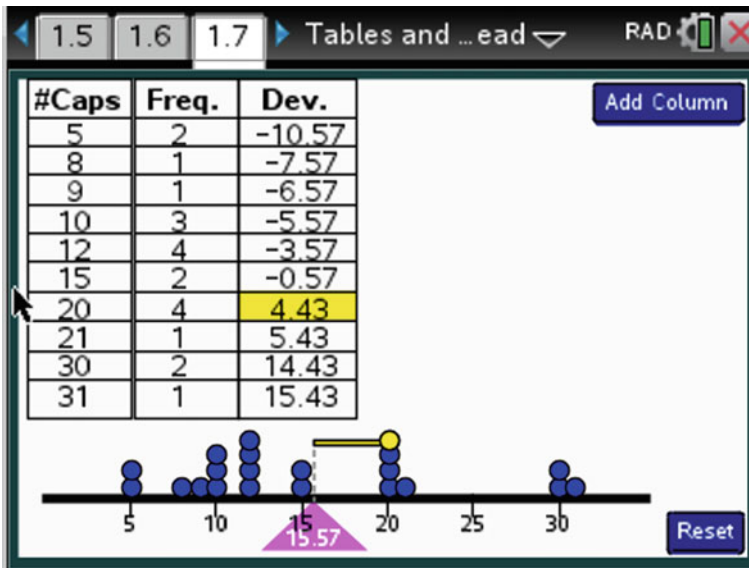


Fig. 6.19 Deviations from the mean

6.4.3 Random Behavior

To most people, “random” events in their lives can be those that are surprising, due to luck or fate, not repeatable or happen just due to “chance” (Batanero 2015). Thus, the natural language learners bring to developing a concept image of randomness is often in conflict with the formal concept definition itself. This can seriously impede the learning of a formal notion of randomness. Students having such a potential conflict in their concept image may be comfortable with their own interpretations of randomness and simply regard the formal theory as not realistic and superfluous (Tall and Vinner 1981). Furthermore, students are bothered by the notion of predicting with some certainty the behavior of a distribution but being unable to predict a specific outcome (Konold 1989). Some believe that it is not possible to apply mathematical methods (statistics) to study random phenomena, because of their unpredictability. Some also believe they can predict or control the outcomes in a random process (Langer 1975).

Given the complexity of building concept images that will enable students to confront their intuitive notions about a random event and align them with the meaning used in statistics, the learning experiences in which students engage need to be carefully designed. Batanero (2015) recommends one possible sequence. First, students learn to discriminate certain, possible and impossible events in different contexts, using the language of chance, and compare an analysis of the structure of an experiment with the frequencies of data collected from repeated experiments to estimate probability. In a second stage, students should move to the study of materials lacking

symmetry properties (e.g., spinners with unequal areas, thumbtacks), where they can only estimate probability from frequencies. The next stage is to investigate real data available from the media, Internet, government or other sources (e.g., sports, demographic, or social phenomena). Finally, students simulate simple situations where the essential features of the situation are modeled by the model used in the simulation and irrelevant properties are disregarded.

Aligned with this framework, the *BCPS* activities introduce the notion of probability using a game where students choose which of two options (i.e., odd, prime) is more likely to occur in drawing ten cards each with a number from one to 10. They have opportunities to play the same game several times and to figure out strategies for winning (the number of successes over the total number of outcomes), giving them experiences that can lead to the creation of a mental structure for estimating probabilities when it is possible to list the outcomes. The technology allows students to simulate the probability using the relative frequencies of a long sequence of drawing cards. The next step is to contrast this situation, where the theoretical outcomes are clear, to a situation where nothing is known about the probability of an outcome (getting a blue chip in drawing a chip from a bag with an unknown number of white and blue chips), using long run relative frequencies to estimate the probability of an outcome (i.e., blue chip). Students generate many repetitions of the experiment, formulate questions or predictions about the trend in the outcomes, collect and analyze data to test their conjectures, and justify their conclusions based on these data. This approach allows students to visualize randomness as a dynamic process in contrast to a printed copy of a random sequence that seems to lose the essence of what random means (Johnston-Wilder and Pratt 2007). The typical sequence of results obtained through repetition lacks a pattern (Fig. 6.20) at the onset. However, “In this apparent disorder, a multitude of global regularities can be discovered, the most obvious being the stabilization of the relative frequencies of each possible result” (Batanero 2015) (see Fig. 6.21).

Students learn that streaks and clusters can appear in a sequence of random outcomes. Technology can be used to create situations involving a cognitive dissonance to help students change their ways of thinking about the concept. In the activity *Choosing Random Samples* students draw names from a hat to identify four students out of 28 to hand in their homework on a given day (Fig. 6.22) (supporting Oehrtman’s third feature of instructional activities (2008)—experience the concept in a variety of situations). Students believe that random behavior somehow balances out in the short run, and once you have been selected you will no longer be called on (Fischbein and Schnarch 1997; Jones et al. 2007; Konold 1989). Simulation allows the process to be repeated many times, and students soon recognize that by chance, a random selection will typically have several students chosen two or even three times in a five-day week. Simulation in a real context can help students establish a better understanding of the nature of randomness. This pattern of random behavior, in the short-term unpredictable but in the long-term stable, is revisited in generating distributions of sample statistics. For random samples selected from a population, students can observe that medians and means computed from random samples will vary from

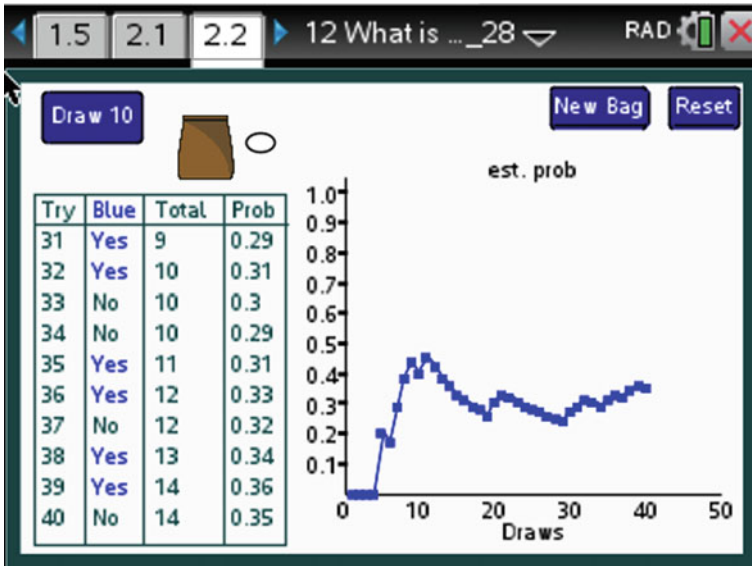


Fig. 6.20 Initial variability

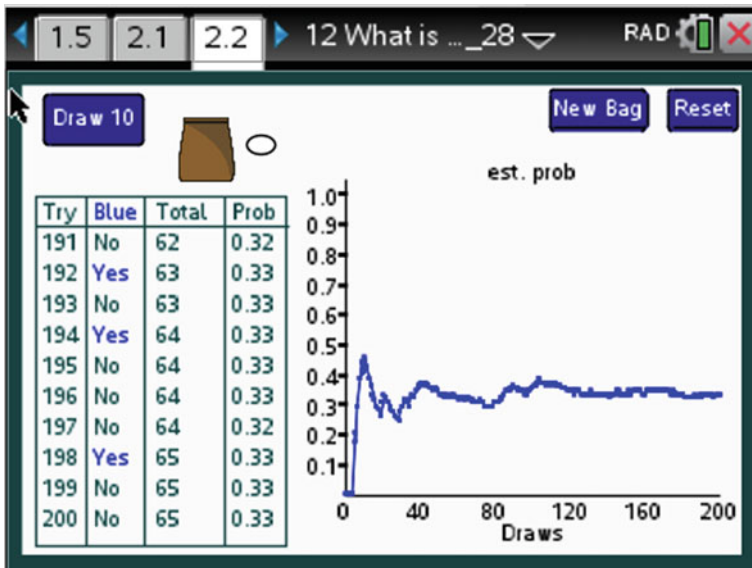


Fig. 6.21 Relative frequency stabilizes

sample to sample and that making informed decisions based on such sample statistics requires some knowledge of the amount of variation to expect.

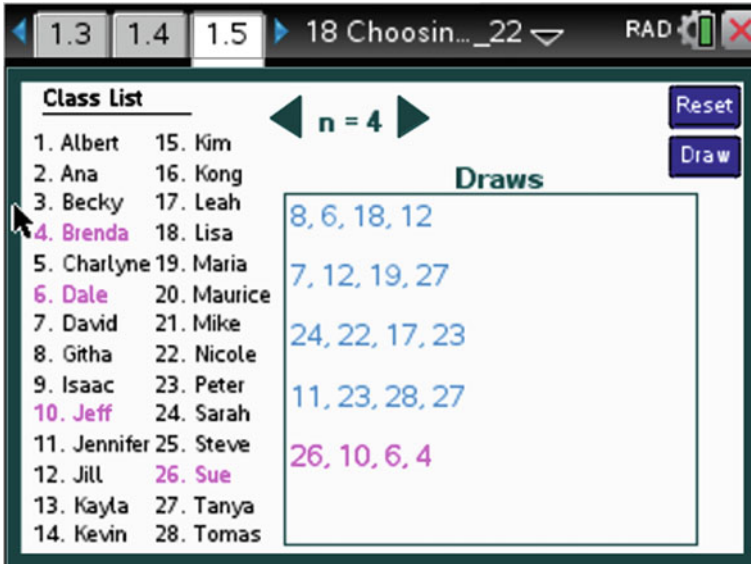


Fig. 6.22 Randomly chosen frequency stabilizes

6.4.4 Sampling Distributions

Students often confuse the three types of distributions related to sampling: distribution of a population, the distribution of a sample from that population, and the sampling distribution of a sample statistic (Wild 2006). In *Samples and Proportions* the notion of distribution is extended from considering the distribution of a population itself and the distribution of a sample from that population to a third kind of distribution, a sampling distribution of statistics calculated from the samples taken from the population (Fig. 6.23).

Students generate many different simulated sampling distributions for a given sample size of the proportion of females from a population that is known to be half female. They discover that each of these distributions seems to be mound shaped and symmetric, centered on the expected value with a consistent range for the number of females in the sample over repeated simulations. A subtle but critical point for learners is that a shift from counts to proportions allows comparison of distributions with different sample sizes and opens up opportunities to think about what is invariant and what is not as the sample size changes and why. The mental image here is highly dependent on noticing that the distinguishing feature is the labeling of the axes. Students can observe that for a sample of a given size, the simulated distribution of the number of females in a sample from a population that is 30% female visually overlaps with the sampling distribution of the number of females in a sample from a population that is 50% female (Fig. 6.24). This leads informally to the concept of margin of error.

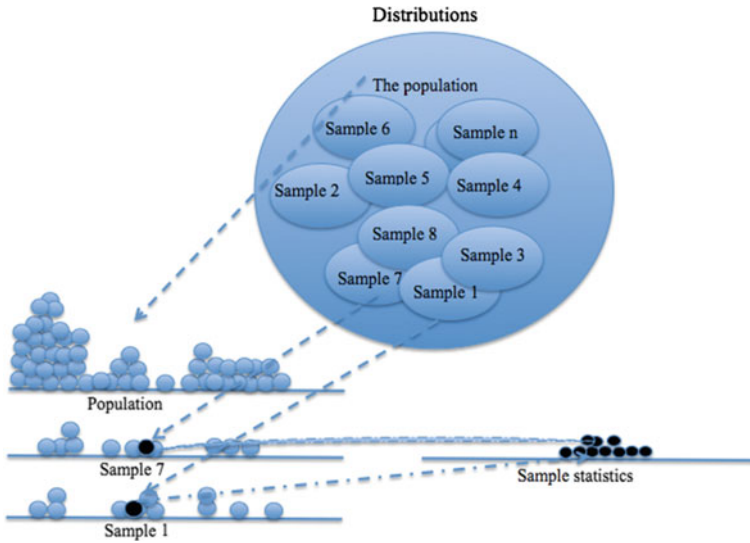


Fig. 6.23 Concept images of three related but distinct notions of distribution

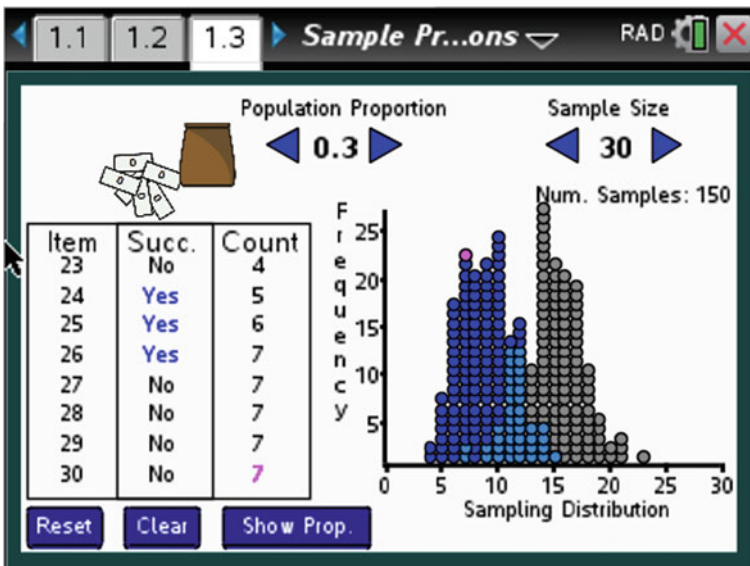


Fig. 6.24 Comparing simulated distributions from two populations

The discussion above described several of the 24 different activities, each addressing particular concepts typically in an introductory statistics course at the school level and as outlined in the CCSS progressions for Statistics and Probability (2011). The files are accompanied by supporting materials that include (1) a description of the

statistical thinking that underlies the file; (2) a description of the file and how to use it; (3) possible mathematical objectives for student learning; (4) sample questions for student investigation; and (5) a set of typical assessment tasks. The activities have been developed for use on a TI© Nspire platform (iPad app, computer software or handheld) and can be downloaded at no cost from the *Building Concepts* website (<https://education.ti.com/en/building-concepts/activities/statistics>).

6.5 Implementation

The interactive documents were used in a semester long statistics course for elementary preservice students. Students had their own computers, and they accessed the materials using the TI Nspire software, although they did use other statistical software packages towards the end of the course. The goals of the course were to enable students to be literate consumers of statistical data related to education and to give them tools and strategies for their own teaching. Student learning experiences were designed with attention to the action/consequence cycle described in Sect. 3.1, an action or activating event, critical reflection, reflective discourse and taking actions based on the new perspective. The students typically worked in pairs or groups on predesigned tasks using the technology to investigate situations, make and test conjectures, usually comparing their results with classmates and engaging in student led discussions on their thinking about the ideas.

6.5.1 Background

The students were sophomores or juniors in the elementary teacher preparation program at a large Midwestern university. They all had selected a mathematics emphasis for their certification (and had taken calculus, which enabled them to interpret the point of inflection on a relatively normal distribution as approximately one standard deviation from the mean); 24 had no prior experience with statistics; three had taken an Advanced Placement statistics course in high school and two had taken a university statistics course.

In keeping with the GAISE framework (Franklin et al. 2007) and the focus of the research, one emphasis in the course was on variability. The next section briefly describes how the interactive documents and action/consequence cycle played out with respect to helping students understand the role of variability in statistical reasoning.

6.5.2 *Instruction*

An “activating event” was an activity or question that engaged students’ curiosity and lead to an investigation of a statistical concept. In the second week of the course, students were asked: How long did it typically take a student in our class to get to campus today? Students made conjectures, then lined up across the classroom according to their times (without talking to make things interesting). The class reflected on the visual representation they had formed; eventually realizing they needed to regroup as they had neglected to consider scale and had just ordered themselves. The distribution of their times now had several clusters and one clear outlier at 90 min. The distribution was reproduced on the board, and the class considered the question: How would you describe the “typical time”. This led to language like “a center cluster”, which motivated a discussion of median, interquartile range and how these ideas would be useful in identifying the typical time to campus. (It turned out to be from 5 to 15 min for half of the students with the median at 8 min.) The outlier was described as surprising.

Cycling through the process, students applied the questions “what is typical” and “what would be surprising” in a variety of situations and new experiences, working with their classmates in randomly assigned groups on tasks creating different graphical representations (action/consequence) and considering the variability in each (reflection). As in the literature, they initially confused variability with range: “[A has] Most in variability: there are many observations on pairs of shoes that people own covering a wide range.” But the majority of students were able to correctly make statements such as: “In variability the height of the peaks don’t matter. Additionally, we are only really looking at the center and how the graph looks around it.”

In the application part of the cycle, students began to use the concept of variability in meaningful ways. For example, looking at the achievement of fourth grade students in science, one student wrote, “An interesting thing about scores from 2015, as can be seen in Fig. 6.2, is that there is an outlier, a state with an average score of 140. This is interesting because while this score is an outlier in the 2015 data, this is not an outlier in 2005, in fact it is part of the lower quartile range. This indicates that in 2015 a score that low would be somewhat unusual, because higher scores are being achieved in science by all of the other states.” They did however continue to struggle with language: “When comparing the western states’ funding to the eastern state’s [sec] funding, the eastern states have a larger range in terms of IQR.”

In a similar fashion, activities such as the soccer tournament described above motivated the use of $\text{mean} \pm \text{MAD}$ (mean absolute deviation) and eventually the standard deviation. Students used simulation to establish what is typically the pattern for a sampling distribution for a given population proportion and sample size. Individually repeating the process over and over (action) and comparing distributions across the class (consequence) gave students the opportunity for critical reflection and to recognize the distribution will always be mound shaped and symmetric with the mean and median around the expected value and one standard deviation at approximately the point of inflection if a smooth curve were drawn over the simulated sampling

distribution. Students noticed the variability in number of successes for a large number of samples of the same size is typically bounded as they simulated the event many times; for example, for a population proportion of 0.5 and sample size 100, the number of successes will rarely be less than 35 or more than 65. “Is it surprising?” led to the activating question—just what does it mean to be surprising? What if for the example, an observed outcome was 34. How do we communicate the notion of surprise at such an observation to other classmates? The discussion and reflection on how to quantify or find a measure for surprising led to the notion of significance.

6.5.3 Initial Results

An initial analysis of some of the data suggests that students for the most part have a relatively solid grasp of variability. For example, the variability around student scores on a state achievement test was given as margins of error (Fig. 6.25). When asked on the final exam, for which student, A, B, or C was the margin of error most problematic, 48% were able to correctly identify student B and 28% answered choice A or C with appropriate reasoning.

Some had the correct answer but incorrect or unclear reasoning; e.g., “This is because with the margin of error, there are lower possible answers than the other students that students with that scale score could have obtained.”

In comparing the standard deviations for the length of time males and females could stand on one foot, 48% of the students associated the standard deviation with the mean but they continued to struggle with precision of language (“The difference in standard deviation between males and females indicates that females are more clustered around the mean.”), while 45% described the variability in general terms without reference to the mean. When asked what image comes to mind when you think about variability, 21% of the students connected variability to the spread

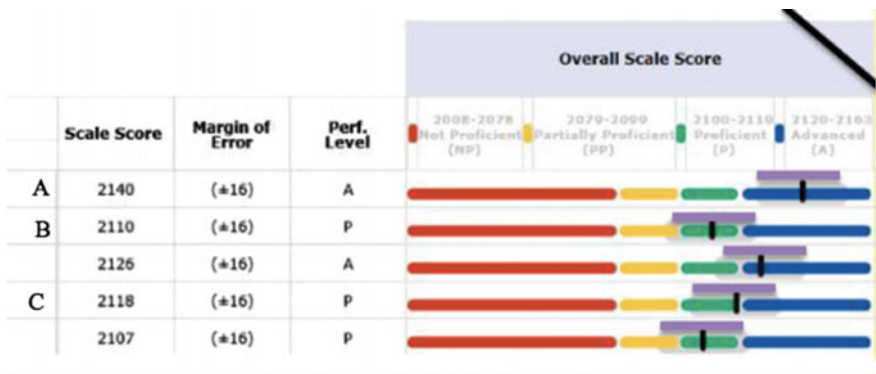


Fig. 6.25 Student achievement results

around the mean or median, 17% gave a general description such as “apartness”, “differentness”, while 31% gave a measure (MAD, IQR, standard deviation).

In comparing the achievement scores of boys and girls, a response such as the following was typical: “...we find that 64.4% of the time this could occur just by chance. We can use this to answer our question and say that grade 5 girls weren’t more likely to score below basic than grade 5 boys because our results could have occurred just by chance. They’re not statistically significant so we can’t say that either gender is more likely to do worse than the other gender based on these results.”

6.6 Conclusions, Future Directions and Research Recommendations

The study was purely observational, with no comparison group or controls for factors such as prior knowledge (although the class as a whole came with little exposure to statistics), which limits any conclusions that can be made. Initial results do seem to suggest the approach has potential for supporting the development of student understanding of variability in multiple statistical contexts. However, the research connecting concept images to visualization to dynamic interactive technology is sparse and a space where much work remains. Some possible questions include:

- What aspects of pedagogy are significant in the use of visualization through dynamic interactive technology in learning mathematics?
- How can teachers help learners use dynamic interactive technology to make connections between visual and symbolic representations of statistical ideas?
- How might dynamic interactive technology be harnessed to promote statistical abstraction and generalization?
- How do visual aspects of interactive dynamic technology change the dynamics of the learning of statistics?

In 1997, Ben-Zvi and Friedlander noted that technology for teaching and learning has evolved over the years, progressively allowing the work to shift to a higher cognitive level enabling a focus on planning and anticipating results rather than on carrying out procedures. Since then technology has provided powerful new ways to assist students in exploring and thinking about statistical ideas, allowing students to focus on interpretation of results and understanding concepts rather than on computational mechanics. While visualizing mathematical concepts has been considered important in developing understanding of these concepts, dynamic interactive technology provides opportunities for students to build more robust conceptual images—to develop video images in their minds as they consider what a concept means in a given context. The *Building Concepts* work thus far suggests that interactive dynamic technology affords students opportunities to build concept images of statistical concepts that align with desirable conceptions of those concepts. The carefully designed action/consequence documents seem to have the potential to be useful tools in providing students with the experiences they need to develop the robust

concept images of core statistical concepts that will enable them to use statistics as a way to reason and make decisions in the face of uncertainty.

References

- Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning*, 7(3), 245–274.
- Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about boxplots? In G. Burrill & M. Camden (Eds.), *Curriculum development in statistics education: International association for statistics education 2004 roundtable* (pp. 163–173). Voorburg, the Netherlands: International Statistics Institute.
- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distributions. 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., & Van Eerde, H. (2014). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbals, C. Knipping, & N. Presmeg (Eds.), *Doing qualitative research: Methodology and methods in mathematics education* (pp. 429–466). New York: Springer.
- Batanero, C. (2015). Understanding randomness: Challenges for research and teaching. In K. Kriner (Ed.), *Proceedings of the ninth congress of the European Society for Research in Mathematics Education* (pp. 34–49).
- Baumgartner, L. M. (2001). An update on transformational learning. In S. B. Merriam (Ed.), *New directions for adult and continuing education*, no. 89 (pp. 15–24). San Francisco, CA: Jossey-Bass.
- 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., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45(1–3), 35–65.
- Ben-Zvi, D., & Friedlander, A. (1997). Statistical thinking in a technological environment. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 45–55). 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. 643–690). Springer.
- Building Concepts: Statistics and Probability. (2016). Texas Instruments Education Technology. <http://education.ti.com/en/us/home>.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–144.
- Breen, C. (1997). Exploring imagery in P, M and E. In E. Pehkonen (Ed.), *Proceedings of the 21st PME International Conference*, 2, 97–104.
- Burrill, G. (2014). Tools for learning statistics: Fundamental ideas in statistics and the role of technology. In *Mit Werkzeugen Mathematik und Stochastik lernen* [Using Tools for Learning Mathematics and Statistics], (pp. 153–162). Springer Fachmedien Wiesbaden.
- 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 (2). <http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art2>.
- Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, 21(1), 1–78.

- Common Core State Standards. (2010). *College and career standards for mathematics*. Council of Chief State School Officers (CCSSO) and National Governor's Association (NGA).
- Cranton, P. (2002). Teaching for transformation. In J. M. Ross-Gordon (Ed.), *New directions for adult and continuing education, no. 93* (pp. 63–71). San Francisco, CA: Jossey-Bass.
- 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*, [Online] 7(3). (www.amstat.org/publications/jse/secure/v7n3/delmas.cfm).
- delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4(1), 55–82.
- Dreyfus, T. (1991). On the status of visual reasoning in mathematics and mathematics education. In F. Furinghetti (Ed.), *Proceedings of the 15th PME International Conference, 1*, 33–48.
- 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.
- Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgments in children and adolescents. *Educational Studies in Mathematics*, 22(6), 523–549.
- Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, 28(1), 96–105.
- Free Dictionary. <http://www.thefreedictionary.com/statistical+distribution>.
- Friel, S. (1998). Teaching statistics: What's average? In L. J. Morrow (Ed.), *The teaching and learning of algorithms in school mathematics* (pp. 208–217). Reston, VA: National Council of Teachers of Mathematics.
- Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, 4(1), 92–99.
- Groth, R., & Bergner, J. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning*, 8(1), 37–63.
- Gould, R. (2011). *Statistics and the modern student*. Department of statistics papers. Department of Statistics, University of California Los Angeles.
- Guin, D., & Trouche, L. (1999). The complex process of converting tools into mathematical instruments: The case of calculators. *International Journal of Computers for Mathematical Learning*, 3(3), 195–227.
- Hancock, C., Kaput, J., & Goldsmith, L. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337–364.
- Hodgson, T. (1996). The effects of hands-on activities on students' understanding of selected statistical concepts. In E. Jakbowski, D. Watkins & H. Biske (Eds.), *Proceedings of the eighteenth annual meeting of the North American chapter of the international group for the psychology of mathematics education* (pp. 241–246).
- Johnston-Wilder, P., & Pratt, D. (2007). Developing stochastic thinking. In R. Biehler, M. Meletiou, M. Ottaviani & D. Pratt (Eds.), *A working group report of CERME 5* (pp. 742–751).
- Jones, G., Langrall, C., & Mooney, E. (2007). Research in probability: Responding to classroom realities. In F. K. Lester (Ed.), *The second handbook of research on mathematics* (pp. 909–956). Reston, VA: National Council of Teachers of Mathematics (NCTM).
- Kader, G., & Mamer, J. (2008). Contemporary curricular issues: Statistics in the middle school: Understanding center and spread. *Mathematics Teaching in the Middle School*, 14(1), 38–43.
- Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development*. New Jersey: Prentice-Hall.
- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6(1), 59–98.
- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32(2), 311–328.
- Learning Progressions for the Common Core Standards in Mathematics: 6–8 Progression probability and statistics (Draft). (2011). Common Core State Standards Writing Team.

- Mathematics Education of Teachers II. (2012). Conference Board of the Mathematical Sciences. Providence, RI and Washington, DC: American Mathematical Society and Mathematical Association of America.
- Mathews, D., & Clark, J. (2003). *Successful students' conceptions of mean, standard deviation and the central limit theorem*. Unpublished paper.
- Mezirow, J. (1997). Transformative learning: Theory to practice. In P. Cranton (Ed.), *New directions for adult and continuing education, no. 74*. (pp. 5–12). San Francisco, CA: Jossey-Bass.
- Mezirow, J. (2000). Learning to think like an adult: Core concepts of transformation theory. In J. Mezirow & Associates (Eds.), *Learning as transformation: Critical perspectives on a theory in progress* (pp. 3–34). San Francisco, CA: Jossey-Bass.
- Michael, J., & Modell, H. (2003). *Active learning in secondary and college science classrooms: A working model of helping the learner to learn*. Mahwah, NJ: Erlbaum.
- Mokros, J., & Russell, S. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26(1), 20–39.
- National Research Council. (1999). In J. Bransford, A. Brown & R. Cocking (Eds.), *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Oehrtman, M. (2008). Layers of abstraction: Theory and design for the instruction of limit concepts. In M. Carlson & C. Rasmussen (Eds.), *Making the connection: Research and teaching in undergraduate mathematics education* (pp. 1–21).
- Peirce, C. S. (1932). In C. Hartshorne & P. Weiss (Eds.), *Collected papers of Charles Sanders Peirce 1931–1958*. Cambridge, MA: Harvard University Press.
- Peirce, C. S. (1998). *The essential Peirce: Selected philosophical writings, Vol. 2 (1893–1913)*. The Peirce Edition Project. Bloomington, Indiana: Indiana University Press.
- Piaget, J. (1970). *Structuralism*. (C. Maschler, Trans.). New York: Basic Books, Inc.
- Piaget, J. (1985). *The equilibration of cognitive structures* (T. Brown & K. J. Thampy, Trans.). Chicago: The University of Chicago Press.
- Posner, G., Strike, K., Hewson, P., & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
- Presmeg, N. C. (1994). The role of visually mediated processes in classroom mathematics. *Zentralblatt für Didaktik der Mathematik: International Reviews on Mathematics Education*, 26(4), 114–117.
- Sacristan, A., Calder, N., Rojano, T., Santos-Trigo, M., Friedlander, A., & Meissner, H. (2010). The influence and shaping of digital technologies on the learning—and learning trajectories—of mathematical concepts. In C. Hoyles & J. Lagrange (Eds.), *Mathematics education and technology—Rethinking the mathematics education and technology—Rethinking the terrain: The 17th ICMI Study* (pp. 179–226). New York, NY: Springer.
- Shaughnessy, J., Watson, J., Moritz, J., & Reading, C. (1999). School mathematics students' acknowledgement of statistical variation. In C. Maher (Chair), *There's more to life than centers. Presession Research Symposium, 77th Annual National Council of Teachers of Mathematics Conference*, San Francisco, CA.
- Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2), 151–169.
- Taylor, E. W. (2007). An update of transformative learning theory: A critical review of the empirical research (1999–2005). *International Journal of Lifelong Education*, 26(2), 173–191.
- Thompson, P. (2002). Didactic objects and didactic models in radical constructivism. In K. Gravemeijer, R. Lehrer, B. V. Oers & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 191–212). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Watson, J., & Fitzallen, N. (2016). Statistical software and mathematics education: Affordances for learning. In L. D. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (3rd ed., pp. 563–594). New York, NY: Routledge.
- Watson, J., & Moritz, J. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), 44–70.

- Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10–26. <http://www.stat.auckland.ac.nz/serj>.
- Zehavi, N., & Mann, G. (2003). Task design in a CAS environment: Introducing (in)equations. In J. Fey, A. Couco, C. Kieran, L. McCullin, & R. Zbiek (Eds.), *Computer algebra systems in secondary school mathematics education* (pp. 173–191). Reston, VA: National Council of Teachers of Mathematics.
- Zull, J. (2002). *The art of changing the brain: Enriching the practice of teaching by exploring the biology of learning*. Alexandria VA: Association for Supervision and Curriculum Development.

Chapter 7

Informal Inferential Reasoning and the Social: Understanding Students' Informal Inferences Through an Inferentialist Epistemology



Maike Schindler and Abdel Seidouvy

Abstract Informal statistical inference and informal inferential reasoning (IIR) are increasingly gaining significance in statistics education research. What has not sufficiently been dealt with in previous research is the social nature of students' informal inferences. This chapter presents results from a study investigating seventh grade students' IIR in an experiment with paper helicopters. It focuses on students' reasoning on the best rotor blade length, addressing statistical correlation. We study how students draw inferences when working in a group; and how their inferences emerge socially in their IIR. For grasping the reasoning's social nature and its normativity, we use inferentialism as background theory. The results illustrate how students' informal inferences are socially negotiated in the group, how students' perceived norms influence IIR, and what roles statistical concepts play in students' IIR.

Keywords Generalization from data · Inferentialism
Informal inferential reasoning (IIR) · Informal statistical inference (ISI)
Informal statistical reasoning · Norms · Social

7.1 Introduction

Mathematical reasoning is social through and through. (Roth 2016, p. 126)

The influence and the role of data for prediction and decision making (see Bakker et al. 2009; Watson 2001) makes generalization from data one of the most influential topics in statistics (Pratt and Ainley 2008). As such, teaching and learning about generalization and inference become key concerns in statistics education. Generalization

M. Schindler (✉)

Faculty of Human Sciences, University of Cologne, Cologne, Germany
e-mail: maike.schindler@uni-koeln.de

A. Seidouvy

School of Science and Technology, Örebro University, Örebro, Sweden
e-mail: abdel.seidouvy@oru.se

© Springer Nature Switzerland AG 2019

G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_7

153

is crucial to inference and researchers have reported students' difficulties in making generalizations (Ben-Zvi and Arcavi 2001). Alongside, there has been an increasingly strong call for statistics education to take into account informal inference (Bakker and Derry 2011; Dierdorff et al. 2011) because an informal approach to statistical inference is necessary in the early years when formal inferential ideas and techniques are beyond young learners' reach (Meletioui-Mavrotheris and Papanastasiou 2015). Informal inference takes into account aspects such as students' prior experiences and their knowledge of real-life contexts. Such previous statistical knowledge creates an arena for students' reasoning in making sense of the data in giving explanations (Gil and Ben-Zvi 2011).

Thus, theories that bridge exploratory data analysis and formal statistical inference (ibid.) have come into focus, especially Informal Statistical Inference (ISI) and Informal Inferential Reasoning (IIR). Research has so far largely focused on portraying ISI and IIR not only as alternative to formal, but also as a tool to shed light on important aspects of statistical reasoning (e.g., Makar and Rubin 2009, 2018; Zieffler et al. 2008). Studies focused, for instance, on the role of the context in developing the reasoning about ISI (Gil and Ben-Zvi 2011; Makar and Ben-Zvi 2011; Pfannkuch 2011). As an example, Gil and Ben-Zvi (2011) showed how a technology-enhanced learning environment can promote students' IIR. There is also a considerable number of scholars attempting to clarify what researchers mean by declaring reasoning as *informal* (Makar and Rubin 2009; Rossman 2008; Pfannkuch 2006).

Even though IIR and ISI are increasingly studied in statistics education research, the question of how students' informal inferences from data emerge *socially* has not sufficiently been addressed. Recent research has predominantly focused on the centrality of data in the generalization process and in making informal inferences (Makar and Ben-Zvi 2011; Makar and Rubin 2009). However, we see that students' generalization from data can hardly be conceived in its entire scope if the social is disregarded or understood only as a context in which reasoning takes place (e.g., Pfannkuch 2011). We share with Roth (2016) the idea that mathematical (here: informal statistical) reasoning emerges socially and we think that investigating students' IIR needs analytical approaches that cope with the question of how students' ISI *emerge* in social situations. Such focus can shed light on the social nature of students' generalizations from data and beyond data.

The purpose of this chapter is, thus, to investigate how students' IIR takes place in social situations and emerges socially. Seventh grade students' IIR was under investigation in students' group work using the philosophical theory of inferentialism (Brandt 1994, 2000). Inferentialism understands reasoning as fundamentally social and normative. The analysis illustrates how IIR emerges socially and is influenced by norms; it illustrates students' roles in the course of their reasoning, as well as the statistical concepts' roles in this process.

7.2 Theoretical Background

7.2.1 *ISI and IIR*

ISI was introduced to capture young learners' statistical inferences before their introduction to formal statistical techniques such as calculating p -values and confidence intervals. ISI is supposed to support and assist students to gain deep understanding of the purpose and utility of data (Ainley et al. 2006) and how data can become an integral part of contextual meaning making (Makar and Rubin 2009). Therefore, an important endeavor in statistics education research has been to broaden the concept of inference (Bakker and Derry 2011; Makar and Rubin 2018; Rossman 2008). However, Pratt and Ailey (2008) have pointed out that what counts as "informal inference" is not easy to determine: "[W]hat is informal could depend on the nature of the inferential tasks being studied, on the complexity of the statistical and probabilistic concepts involved, on the educational stage, and on other factors" (p. 3). In this study, we adopt the view of Bakker and Derry (2011) and Bakker et al. (2006) that broadens the meaning of statistical inference to enclose more informal ways of reasoning and to include judgment based on contextual knowledge (Makar et al. 2011). Briefly speaking, in this study we draw on Makar and Rubin's (2009) framework for thinking about ISI, based on a broader reasoning process that takes into account human judgment of the statistical context. IIR is the reasoning process leading to and underpinning ISI (Makar et al. 2011). Gil and Ben-Zvi (2011) argue that IIR could bridge the gap between exploratory data analysis and formal statistical inference. Following Rubin et al. (2006), IIR can be understood as statistical reasoning that takes into account several dimensions such as "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" (Gil and Ben Zvi 2011, p. 88). In this study, we adopt a working definition proposed by Zieffler et al. (2008) who describe IIR as "the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples" (p. 44).

Makar and Rubin (2009) developed the theory about ISI. They identified three essential key principles: (1) *generalization*, including predictions, parameter estimates, and conclusions, that extend *beyond describing the given data*; (2) the use of data *as evidence* for those generalizations; and (3) employment of probabilistic language in describing the generalization, including informal reference to levels of certainty about the conclusions drawn.

The first key feature, *generalization beyond data* signifies making inferences outside the data at hand, outside a given sample. Unlike generalization in mathematics, which tends to be deterministic, generalization in statistics and probability deals with uncertainty (Burgess 2006; Groth 2013). Put simply, making statements about a population based on a sample requires taking into account the variation of and uncertainty within data. Variation and uncertainty do not necessarily facilitate generalization from sample to population—in fact, variation and uncertainty make

generalization elusive. Alongside, research has shown that students' tendency to treat data as individual values may prevent them from experiencing data as a global entity about which generalizations can be made (Ben-Zvi and Arcavi 2001).

The second key feature is the *use of data (sample) as evidence* for generalization. What is considered as evidence is what can serve as means to support a claim (generalization) in a given context. As such, evidence is contextual. Connecting data to the context can help students in making sound generalizations (Gil and Ben-Zvi 2011; Makar and Ben-Zvi 2011; Pfannkuch 2011). Evidence is also social in that it has to be accepted in a given community. Data as evidence refers to numerical, observational, descriptive, or even unrecorded evidence (information) that is accepted by the statistics research community (Makar and Rubin 2009).

The third feature in Makar and Rubin's characterization of statistical inference is the use of *probabilistic language in describing the generalization*. Making a statement about a population based on the sample must deal with uncertainty. In formal statistical inference, this uncertainty can be quantified using statistical techniques, for instance in form of a certain percentage in a given confidence interval. In IIR, uncertainty is expressed in a broader sense, in students' informal ways describing with their own words, without formal statistical calculation, using words such as "maybe", "it could be", etc. (Makar et al. 2011).

7.2.2 Reasoning and the Social from an Inferentialist Perspective

As a background theory, inferentialism determines the theoretical foundations our study relies on. As Noorloos et al. (2017) have pointed out, inferentialism holds the potential to overcome some of the philosophical problems of socioconstructivism, such as the unsatisfactory resolution of the social-individual dichotomy.¹

Inferentialism is a semantic theory rooted in pragmatics. It is based on philosophical ideas offered by for example, Kant, Hegel, Frege, Wittgenstein, and Heidegger (Brandom 1994, 2000). Inferentialism as background theory provides what Roth (2016) calls a sociogenetic perspective to students' reasoning. Inferentialism does not understand the social as a context for the individual: It rather assumes that reasoning is social in the way it exists in students' discussions (see Schindler et al. 2017).

¹A comprehensive description of inferentialism, its differences from constructivism, and its strengths can be found in Noorloos et al. (2017).

7.2.2.1 Philosophical Background: Language Games and the Game of Giving and Asking for Reasons (GoGAR)

One of the key ideas that inferentialism draws on is the concept of language game, provided by Wittgenstein (1958) in his late works. With this term, Wittgenstein describes the practice in which participants provide utterances, stating “I shall call the whole, consisting of language and the actions into which it is woven, the ‘language game’” (p. 5). The language game is significantly influenced by societal framework conditions, such as authority relations, duties, and responsibilities (Newen and Schrenk 2012). The very nature of the language game is not determined by grammatical rules, but by the community’s practices and courses of conducts (ibid.). The expression *game* refers to the significance of such habits for making moves (i.e., utterances) in the linguistic practice. In turn, the significance of words in the linguistic game can only be understood when such “rules” of the game (norms, expectations, etc.) are taken into consideration. Wittgenstein uses the metaphor of chess game to introduce the concept of language game: “The question ‘What is a word really?’ is analogous to ‘What is a piece in chess?’” (Wittgenstein 1958, p. 47). In turn, describing the pieces in the chess game via the rules of the chess game, not via their physical properties, is analogous to describing words via the rules of their use in the linguistic practice, not via their properties. However, it is not only the rules of the game that determine the meaning of utterances, but it is also their position in the game. Peregrin (2009) also uses the metaphor of chess to describe this:

Chess once more: Though the pieces have their ‘position-independent’ roles which reflect their ‘force’ (the role of the queen makes the queen a much more powerful piece than the pawn), the significance of pieces for a particular player in a particular position need not always reflect this: there are (rare) positions in which the knight is more useful than the queen (p. 167).

The position of an utterance in a discussion can give the utterance significant power for the ongoing discourse. Brandom (1994, 2000) draws on these ideas when conceptualizing the *game of giving and asking for reason (GoGAR)*—a pivotal concept in inferentialism. Brandom (2001) explains that the significance of utterances relies on the language game, in which persons give reasons and ask for reasons. Utterances represent certain moves in the game (taking up a certain position); a game that consists of giving and asking for reasons. These moves in the language game contain a commitment to a certain content that can serve both as premise and as a conclusion in reasoning and inferring (ibid.).

7.2.2.2 Epistemological Considerations Derived from Inferentialism

Students’ reasoning in the GoGAR. We understand students’ discourse in their statistical inquiry as a GoGAR, in which they make claims, give reasons, ask for reasons, acknowledge claims and reasons, attribute them to others, undertake them themselves, or reject them. The term *game* highlights the significance of certain

(possibly implicit) rules about how to make moves; about how to bring forward claims in the statistics classroom, about how, when and whom to ask for reasons, etc. When generating and manipulating data, students furthermore make actions (e.g., creating a diagram, showing it to others, pointing onto the solution they think is correct). According to Bakker et al. (2017), inferentialism treats judgment and actions democratically without any prior assumption to hierarchy. Accordingly, Brandom (2000) takes reason (judgement) to be the minimal entity one can take responsibility for on the cognitive side, and action as a minimal one can take responsibility for on the practical side. In our inferentialist epistemology, we consider students' GoGAR not only to consist of students' linguistic reasoning, but also of their actions.

Primacy of the social and normativity. As we have pointed out elsewhere with regard to students' *mathematical* reasoning (see Schindler et al. 2017), we assume students' *statistical* reasoning to be primarily social. This is to say that reasoning emerges socially in the processes of making claims, attributing claims to other persons, acknowledging other persons' claims, undertaking them oneself, etc. We even believe students' reasoning to be social when students are, for instance, working solely; because it has a social origin and reflects the social situations that it is derived from as well as perceived norms (about how to draw inferences, how to use words, etc.). We assume students' IIR to be social at the same time as it is statistical: Students' IIR takes place in—implicitly or explicitly—social situations: for example, in the statistics classroom or in out-of-school situations. It reflects the social situations and the norms embedded in these situations. Brandom (1994) states that the GoGAR in this sense is implicitly *normative*: everyone who is involved in the GoGAR judges about whether or not the moves that the other people make are appropriate or not. In students' group work in the statistics classroom, the appropriateness of students' utterances is considered based on, among others, social, statistical, and mathematical norms in the statistical classroom (e.g., didactical contracts, rules of behavior, etc.). When students base their generalizations beyond the data on the mean value as concept, this may, for example, reflect the perceived norm in the classroom that the teacher views the mean value as an appropriate reason and important concept. It is possible that the same students draw on other concepts (such as a diagram, or extreme value) in other social constellations, for instance, when they work in groups or in out-of-school situations. In our inferentialist epistemology, we argue that students' IIR cannot be isolated from these social and normative factors. We rather assume that students' IIR is social and—in the above-mentioned sense—normative to the core.

Meaning of concepts as constituted through their roles in the GoGAR.

Statistical concepts are understood in the roles they play in the language game; in particular, in students' reasoning and inferences. Bakker and Derry (2011) claim, "statistical concepts such as mean, variation, distribution, and sample should be understood in terms of their role in reasoning, i.e., in terms of the commitments entailed by their use" (p. 11). We think that the significance of a concept for students can only be understood if its role in students' reasoning is taken into consideration. We focus on how the concept is used in IIR and what significance it has. In our

inferentialist epistemology, we think that the meaning of concepts consists in the roles that they play in the GoGAR. Their meaning develops through the evolving roles.

7.2.2.3 Research Questions

Based on our inferentialist epistemology, we ask the following research questions: How does students' IIR and—in consequence—their ISI emerge in the GoGAR in a social practice? In particular, how do students socially make generalizations beyond the data, which evidence do they negotiate, and how do they deal with uncertainty? How is their IIR influenced by perceived norms? What roles do statistical concepts have in students IIR? In particular, how do students use statistical concepts as reasons in IIR?

7.3 Method

7.3.1 Design of the Study

7.3.1.1 Preparing for the Empirical Study

We use data from an empirical study carried out in a class with 20 year 7 students (aged 12/13) in a Swedish secondary school. This study incorporated the “*helicopter task*” (see Pratt 1995, see below) taken from Swedish in-service teacher developmental training program called “*matematiklyftet*,” designed to support in-service teachers.

In preparation for the study, we initiated regular contact with the teacher and the class and had several electronic conversations with the teacher before the teaching experiment. During these conversations, the teacher was introduced to the task and the procedures for conducting the teaching experiment, including the students' envisioned learning trajectory, follow-up questions, etc. The researchers additionally had two preliminary meetings with the teacher where they visited the class during mathematics lessons. The teacher, Mrs. Andersson (pseudonym), had worked as an in-service mathematics teacher for over 20 years. She showed enthusiasm to participate in this study. During preliminary interviews, the teacher discussed students' learning background, their prior experiences in the field of statistics, the social climate in the classroom, and the habits in the class of working collaboratively. It appeared that the class was used to work in groups, but not to conduct experiments themselves. According to the teacher's evaluation, the performance level of the students was low as compared to other mathematics/statistics classes.

The teacher reported that the class was previously introduced to descriptive statistics and graphical representations of data, to the use of diagrams, and methods to calculate the mean and median. The students were familiar with the use of tablet

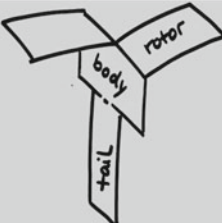
computers across all subjects. However, the class was not used to aggregate aspects of data such as statistical correlation, sample distribution, or variation.

7.3.1.2 The Mathematical Task and the Aims

In order to support students' IIR, we chose a task that was based on experimentation and correlation. In the task, the students were asked to explore paper helicopters; in particular the flying times dependent on the rotor blade lengths (RBLs) (Ainley et al. 2001; Pratt 1995; see Fig. 7.1).

We used the task in a group work setting in order to encourage students to socially discuss different aspects in the data. Our aim was to open up for a social and content related (statistical) GoGAR. We expected the students to deal with uncertainty, because measuring flying times inevitably invites students to cope with uncertainty due to measurement errors. Our hypothesis was that claims dealing with uncertainty would be expressed in probabilistic language beyond data. Generalization (beyond data) was supposed to be addressed when students state or make claims that indicate a correlation between the RBL and flying time. In particular, the students' GoGAR was expected to address correlation based on concepts such as mean value, mode, median, or extreme value (see Mokros and Russell 1995). Alongside, we expected students to make generalization beyond data based on causality (laws of physics), for instance reasoning based on the air resistance, the quality of the paper used, etc. Moreover, we anticipated that the students would make statements following a single data approach (Moritz 2004), for instance, "You *just* have to look at the biggest value. Then you're done." We also conjectured difficulties in representing multiple variables (McClain et al. 2000; Moritz 2000).

Your class was chosen to help a company to design a paper helicopter. Your job is take care of the rotor blades. The rotor blades are supposed to carry the helicopter as long as possible. You are supposed to find the best rotor blade length for the helicopters. The preliminary tests indicated that the lengths of the rotor blades should lie between 3cm and 14cm.



Answer the following questions:

1. Which one do you think is the best rotor blade length? Explain.
2. How sure are you about that?
3. Can you find another way to decide the best rotor blade length?

Fig. 7.1 The paper helicopter task

7.3.1.3 The Design of the Empirical Study

The class dealt with the helicopter task in two lessons with one day between. The focus of the first lesson was to actively engage the students in the data collection. In the first lesson (not in the focus of this chapter), six groups of students each measured five flying times with helicopters within a 3 cm span rotor length (6–8; 9–11; 12–14; see Table 7.1). After the first lesson, the team of researchers collected the data and put them in a table displaying the data of the whole class (Table 7.1). During the second lesson, the students received this data table together with the task.

The focus of the second lesson was students’ IIR. The students worked in new groups (groups 1 & 2 → group A; groups 3 & 4 → group B; groups 5 & 6 → group C). The last activity in the second lesson was a whole class presentation of students’ answers to the task and gathering questions. The teacher organized this discussion. The role of the researchers during these two lessons was limited to observation and video recording.

7.3.2 Data and Data Analysis

We focused on the data of one group (group C, 5 female participants). This group caught our attention during the first lesson when they focused on data generation. Each member had a specific assignment and the group had extensive discussions, which led to rich data for our analysis. During the second lesson, the group was

Table 7.1 Data in the helicopter experiment, gained by the students

| | 1 st | 2 nd | 3 rd | 4 th | 5 th | 6 th | 7 th | 8 th | 9 th | 10 th |
|-------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|------------------|
| 6 cm | 2.50 | 2.30 | 2.50 | 2.30 | 2.40 | 1.90 | 2.59 | 2.94 | 2.56 | 1.83 |
| 7 cm | 1.8 | 1.9 | 2.7 | 2.2 | 2.1 | 2.43 | 3.13 | 2.34 | 2.3 | 2.85 |
| 8 cm | 2.6 | 1.9 | 2.6 | 2.5 | 2 | 2.97 | 2.45 | 2.38 | 3.04 | 2.95 |
| 9 cm | 3.1 | 2.98 | 2.9 | 2.86 | 3.03 | 1.64 | 1.31 | 1.94 | 1.85 | 2.12 |
| 10 cm | 2.2 | 2.6 | 2.43 | 2.82 | 2.38 | 1.65 | 2.47 | 2.27 | 2.18 | 2.24 |
| 11 cm | 2.24 | 2.8 | 2.9 | 2.12 | 2.7 | 1.34 | 1.87 | 1.87 | 1.3 | 2.78 |
| 12 cm | 2.8 | 2.2 | 2.56 | 2.38 | 3.32 | 1.35 | 2.31 | 1.81 | 1.73 | 2.03 |
| 13 cm | 2.56 | 2.23 | 2.46 | 2.28 | 1.73 | 1.68 | 1.63 | 1.81 | 2.38 | 1.98 |
| 14 cm | 4 | 2.98 | 2.95 | 3.08 | 2.08 | 2.33 | 1.86 | 2.16 | 1.68 | 2.36 |

Data gathered by: group 1, group 2, group 3, group 4, group 5, group 6

engaged in the GoGAR with various language moves. We use this group to exemplify how an inferentialist epistemology can contribute to understanding students' IIR in both its statistical and social facets.

The data analysis focuses on the aspects described in Sect. 7.2.2.2. Based on the analysis of students' GoGAR (which includes not only oral reasoning but also students' actions), we investigate students' IRR. In particular, we study how IIR emerges in the social practice, what roles students have in moving the IIR forward, and what roles statistical concepts have in this process.

7.4 Results

The following scenes took place in the second lesson. In this lesson, the teacher initially recalled what had happened during the first lesson. Thereafter, she explained what was planned for the second lesson. The students received a table showing all the data gathered by the students in the first lesson (Table 7.1).

Phase 1: Bringing the diagram into play. Immediately after receiving the task, one of the students, Rose (all names are pseudonyms), showed a tablet computer screen to the other students: It displayed a diagram that she had drawn before this lesson; based on the data that she had gathered in lesson 1 (Fig. 7.2). This diagram incorporated the flying times that Rose's group gathered (group 6, containing also Katie and Lucy, Table 7.1). The other two girls (Anna and Maria) had attended group 5 and gathered other data (Table 7.1).

In this phase, two aspects were distinctive in the GoGAR: First, the other girls claimed and insisted that the flying times of the other group must be integrated in the diagram as well. For instance, Anna insisted, "You have to put in our flying times as well." She called the teacher and complained to her, "She kind of put in only her own values, she did not put in our values." She seemed to draw on the teacher's authority



f.l.t.r.: Maria, Katie, Anna, Lucy, and Rose



Displayed diagram

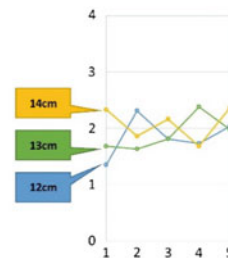


Diagram (reconstructed)

Fig. 7.2 The students' initial diagram (part 1)

in order to make Rose integrate her group's values. The GoGAR indicated that the students saw the necessity to have all their data displayed in the diagram. However, the students did not ask Rose to integrate *all* the data of the table; they rather asked her to integrate "their" values; in this case, the flying times for RBLs 12 to 14 (group 5, see Table 7.1). Their reason for focusing only on "their" data—not on all data—did not become explicit. They possibly misunderstood the task (assuming that they only had to consider their own values), or they (implicitly) perceived that longer RBLs go along with longer flying times (correlation), leading them to only consider the longest RBLs in the experiment. It could also be the case that their own data became subjectively very important to them. In a similar vein, Lehrer and Romberg (1996) and Roth (1996) observed students' engagement in statistical tasks when given the opportunity to generate their data themselves (see also Cobb and McClain 2004).

The second distinctive aspect in this phase was Rose's explanation of the diagram to the other students. She, for instance, explained that "this is the length [pointing to y -axis], this is the time (pointing to x -axis), and these are the trials [pointing to the points in the diagram]", because the other students (especially Maria and Anna) initially claimed that they "do not understand" the diagram and asked for clarification.

Even though the students did not make informal statistical inferences yet in this phase, their GoGAR was important for the following IIR: It layed the foundations for the diagram's role in students' IIR, especially for finding reasons that support their ISI. In this phase, we can also see that the students had different roles in moving the GoGAR forward. Rose came up with the diagram that she had prepared in advance (her actions such as showing the diagram or pointing are crucial). Anna insisted that her values must be integrated as well (and even used the power of the teacher to make Rose integrate the data). Moreover, Maria asked several times for an explanation of the diagram—leading Rose to explain the diagram. The latter gave the other students in this group the opportunity to understand the diagram better; which in turn facilitated the use of the diagram in their IIR. At the end of this phase, Rose started filling the values of group 5 in her diagram.

Phase 2: Informal confidence interval. Directly after the first episode, the students started to reason with the best RBL:

| Turn | Speaker | Statements |
|------|---------|---|
| 01 | Katie | We should decide the best rotor blade length. |
| 02 | Maria | If there had been 15 or 16 then maybe it had gone slower. Slower than 14. |
| 03 | Katie | I think that it is 12 and above uhh...It does not go on forever. |
| 04 | Anna | Not 14 and above, no: 13 and above. |
| 05 | Maria | I think it is 13 and above and the limit is 16 |
| 06 | Anna | Mhm. Like 13 to 16. |
| 07 | Maria | Yeah. There, the best rotor blade lengths may be situated. |

Katie's first statement in this episode (Turn 01) leads the students' GoGAR to focus on the best RBL. Maria's subsequent statement (02) reflects that she is not only taking into account the given data, but extrapolating the data: She makes a prediction based on the data, speculating about the flying times of helicopters with

a RBL of 15 or 16 cm. In particular, she claims that the helicopter would possibly have gone slower if the RBL had been 15 or 16. This indicates that she assumed a correlation in the kind of “the longer the rotor length, the longer the flying time.” However, Katie’s statement (03) indicates that she is aware that this correlation “does not go on forever.” The following moves in the GoGAR (04 to 07) reflect that the other students acknowledge this claim. They claim that it is “12 and above” (03), “13 and above” (e.g., 04): With this idea the students presumably want to express an intuitive confidence interval in the sense that they “have a sense of the reasonable expected variability around the expected value” (Shaughnessy 2006, p. 87). Maria’s utterance, “I think it is 13 and above and the limit is 16” (05) sets the lower and upper boundaries. The question of why the girls seem to be aware that “it does not go on forever”; and, thus, focus on 15/16 as the best RBL; not on 20, for instance, is dealt with at the end of the lesson (see below).

In the subsequent group work—which is not presented in a transcript due to space restrictions—, Anna, Maria, and Katie repeatedly claimed and confirmed that “we believe in 13 till 16.” Additionally, Maria claimed that it may be “15, something like that, I think” and went on reasoning, “it feels like we have to have one rotor length and in that case it is like 15, I think.” The reason for uttering the latter seems to lie in a perceived norm to determine a *single value* as the best RBL. It can be, for example, the formulation of a mathematical task or students’ experiences in the mathematics and statistics classroom (classroom norms) that made her assume such an expectation. Maria’s expression of this perceived norm was influential on the ongoing GoGAR: Katie acknowledged it, stating, “then I think it is 15.” Anna claimed, “but we believe in 15 then”. Maria summed up their preceding GoGAR stating, “but (...) we believe in 13 till 16 (...) one of them but if you have to be exact we can say 15.”²

In this phase, it is—from an analytical point of view—difficult to state which concepts the students drew on that moved the IIR forward. In fact, the question arises of what students’ reasons may have been for assuming the correlation “the longer the rotor length, the longer the flying time.” Why did they assume that the flying time gets longer the longer the rotor blades are? Were they scanning the numbers and, for instance, focusing on 4 s as extreme value, leading to 14 cm as best RBL? Have they implicitly drawn inferences from the diagram that Rose showed them (Fig. 7.2)? At this point, it cannot be verified that the reasons that the students drew on were statistical ones (such as mean value). Nevertheless, we argue that the students’ endeavor, in which they drew on physical properties, in which they seemed to express their uncertainty through an informal confidence interval, drew generalizations beyond the data and used the data at hand, claimed a correlation, and mentioned its limits may be regarded as IIR. This reflects the above-mentioned broadened view on ISI and IIR as proposed by, for example, Gil and Ben-Zvi (2011) and Makar et al. (2011).

Phase 3: Reasoning based on the diagram. After their discussion on the interval, Rose showed the completed diagram (Fig. 7.3):

²Please note that (...) indicates a short pause of speech in the student’s utterance. We did not attempt to quantify the exact length of pauses.



Diagram as displayed on the tablet

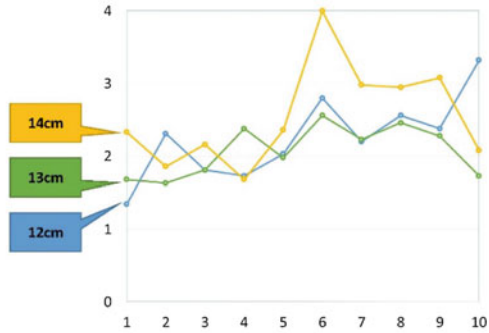


Diagram (reconstructed)

Fig. 7.3 The students' updated diagram (part 2)

| Turn | Speaker | Statements |
|------|---------|--|
| 13 | Rose | Done! Here! (showing the new diagram to the group after integrating the new values) |
| 14 | Maria | What does the yellow stand for? (pointing at the graph on the tablet) Aha, it is the color. |
| 15 | Anna | Blue is 13, right? |
| 16 | Rose | And blue [...] blue is 12 and green is 13. |
| 17 | Maria | Now I can follow you, the blue is ... (looking and pointing at the graph) |
| 18 | Rose | Yes... Now you can SEE ³ which one is the best! |
| 19 | Anna | It is the yellow one. Cheer yellow! |
| 20 | Katie | Mhm. |
| 21 | Anna | Yellow is the best. |
| 22 | Maria | How sure are we? |
| 23 | Katie | Pretty sure! |

The diagram appeared to play an important role for the students when inferring that 14 cm is the best RBL; and it was crucial for their IIR. Rose stated, “You can SEE which one is the best” (18). What the students perceived in the diagram confirmed what they implicitly had assumed beforehand; and it convinced them. Based on the diagram, they were “pretty sure” that 14 cm is the best RBL, indicating that the diagram assisted them to manage their uncertainty in the data. Later on in this lesson, Rose explained to the teacher and to the other classmates that “it is much easier to see the difference between the different rotor lengths [in the diagram, authors’ note]. 14 is highest here [pointing at maximum], but kind of low there [pointing at minimum]. This is because if you drop it a little inclined, or we were two different groups and

³Please note that it was Rose’s emphasis.

may have dropped it in different ways. 12 is quite low compared to 14. And 13 went downwards at the end.”

Phase 4: Reasoning based on the mean value. In the subsequent group work, the students tried to find another way to determine the best RBL (see third question in Fig. 7.1).

| Turn | Speaker | Statements |
|------|-----------------|---|
| 24 | Katie | Did you hear what Rose said? |
| 25 | Anna | No... |
| 26 | Katie | Rose, can you quickly explain it? |
| 27 | Anna | The mean? |
| 28 | Rose | The mean. Then you know approximately where everything is conglomerating. |
| 29 | Katie | If we calculate the mean for every one of these (rows), then we have kind of the mean value. |
| 30 | Maria | This is not the same type as the first one. |
| 31 | Katie/Anna/Rose | No. |
| 32 | Rose | This is now the mean value. |
| 33 | Anna | So, if we calculate the mean value, then we know approximately where all (trials) are conglomerating. |
| 34 | Rose | Mhm. [nodding in agreement] |
| 35 | Anna | Then we have a good calculation. We nailed it. Good job. Now let's calculate the mean. |

In this phase, the students drew on the mean as a statistical concept in their reasoning. The mean had the role as an alternative concept for drawing the inference that 14 cm is the best RBL. Maria, for instance, stated, “this is not the same type as the first one” (30), which is confirmed by three of the other girls. Anna’s final utterance “then we have a good calculation. We nailed it. Good job. Now let’s calculate the mean” (35), indicates a focus on finding a method that made them able to calculate. This may indicate a perceived norm in the classroom to use approaches that draw on calculations. The students furthermore presented this approach to the teacher when she approached their group. They decided that Rose is responsible for calculating the mean values for the three RBLs (12–14 cm). However, neither the results of this calculation nor the mean value in general were further discussed in the GoGAR. The students did not seem to feel the need to confirm the ISI that they had drawn based on the diagram. This indicates that the mean had a subordinate role in the students’ IIR on the best RBL. They seemed to use it as an additional reason that they perceived meets the classroom norms. In the final presentation, Katie summarized, “and then we tried and calculated the mean values for 12, 13, and 14 and then we got approximately the same every time [as compared to the results from the diagram].”

Phase 5: Presenting and summarizing the reasoning. When the students presented their work to their classmates at the end of the lesson, they summarized their work. Here, it also became apparent why they seemed to be aware that the correlation “the longer the rotor blades the longer the flying time” is limited; and thus focused on 15 as best length.

| <i>Turn</i> | <i>Speaker</i> | <i>Statements</i> |
|-------------|----------------|---|
| 36 | Katie | We think that the best rotor length is between 13 and 16 cm. For 14, it worked best, for us. When we tested it with like 20, it got completely too long. |
| 37 | Teacher | So, you tried it out? |
| 38 | Maria | Yes. |
| 39 | Katie | Yes. It didn't work out. The shorter... too short doesn't work either. So it just drops. So, we believe, something between 13 and 16. Well, 15. But if we choose one of our own lengths, then it is 14. |
| 40 | Teacher | Okay. |

In Turns 36 and 39, Katie summarized their reasons for finally choosing 14 as best RBL: They reasoned based on the physicality that the helicopter dropped when it is either too long or too short. Katie additionally expressed the informal confidence interval; and pointed out the distinction between the correlation that the students perceived within the data (“if we choose one of our own lengths, then it is 14”) and the generalization *beyond* the data (“so, we believe, something between 13 and 16. Well, 15.”)

7.5 Discussion

The purpose of this chapter is to investigate students' IIR in social practices (Brandom 1994, 2000). In particular, we inquired into the roles that individual students, different informal and formal statistical concepts, and norms may have on students' IIR.

The group work of five grade 7 students, Maria, Katie, Anna, Lucy, and Rose, illustrates how perceived norms and students' moves in the course of the GoGAR contributed to the emergence and development of IIR. It illustrates that ISI, the resulting end-product of IIR, is consequently the result of a synergy of statistical content and social practice, in which the individual students participate and each have their roles. One student brings, for instance, a diagram into play, another student insists to have more values integrated, and a third student asks for explanation. In sum, we see that students' IIR in the group work and the resulting ISI consist of the moves that the individual students make in the GoGAR. The students take—consciously or not—different roles in moving the GoGAR and the IIR forward. In our study, four out of five students in the group set different impulses that move and guide the IIR. The students contribute to the ISI through actions (e.g., showing the diagram), through claims on the content (e.g., “it does not go on forever”), questions (e.g., “can you explain this?”), or requests (e.g., “you have to put in our values as well”). We observe that normative aspects have a significant influence on the course of students' IIR. The students seem to perceive, for instance, the norm to have a single value as their answer (instead of the informal confidence interval that they believe in), or the norm to find an approach that draws on calculation (the mean value), or to follow the questions listed in the task. The students even appear to make use of normative

aspects consciously. One student draws on the authority of the teacher to make her groupmate integrate her values in the diagram.

Furthermore, the analysis illustrates the roles that the statistical ideas can take in students' IIR. In the group at hand, the *diagram* has a significant role; which can be seen in students' reasoning within the group work as well as in their presentation at the end of the lesson. The students stress the fact that they can "see" the best RBL in the diagram. It seems as if this representation is easier to grasp for them than the mean, which has a subordinate role in the students' GoGAR. Students' problems in their conceptual understanding of the mean (as observed in Gal et al. 1990; Makros and Russell 1995; Pollatsek et al. 1981) may be one reason why they do not use it naturally in their reasoning. In a large part of the GoGAR, the students discuss the best RBL without explicating which concepts they draw on: They probably feel that they share the same view and do not need to make the underlying reasons explicit (Cobb and McClain 2004). Finally, the students discuss the mean value as a means to determine the best RBL. They do not do so spontaneously, but rather react to the prompt in the task. The mean value has the role to confirm what the students' already saw in the diagram. They draw on the mean value in order to "have a good calculation." In fact, the students mainly mention the mean value when communicating with their teacher. It appears as if the students follow a perceived expectation by the teacher to use the mean as a concept; however, they seem to "trust" the diagram in their own reasoning. This illustrates the normative dimension of IIR: The reasons that the students draw on may be different depending on the persons that are involved and their perceived expectations.

Another normative aspect in students' group work is their single value approach. When discussing the best RBL, the students informally appear to discuss confidence intervals (see Shaughnessy 2006); claiming, for instance, that the best length must lie between 13 and 16. This connects to research results by Makar and Rubin (2009), who found that students in their study broadened their predictions to range values in order to improve their predictions. Broadening predictions is a strategy that students use to decrease uncertainty and to avoid being wrong. On the other hand, one student brings the perceived norm into play that one has to provide a precise, single-valued answer, stating "it feels like we have to have one rotor length" (Maria). This has an impact on students' IIR: Their reasoning accordingly focuses on finding a single value. This connects to research in the domain of statistics education acknowledging and describing the impact of norms on students' IIR (e.g., Bakker and Derry 2011; Makar et al. 2011). However, in contrast to Makar et al. (2011), who point out that there exist cognitive and social elements in students' IIR (p. 169), we avoid to differentiate between cognitive and social aspects. We rather see students' utterances as always influenced by norms and, thus, as normative themselves. Besides, the fact that the students in our study frequently refer to what they "believe" in, relates to Makar et al.'s (2011) assertion that beliefs are among the key drivers in students' IIR.

Finally, our analyses and the results confirm that our epistemological theory based on inferentialism allowed us to grasp students' IIR as a social practice. This background theory assumes—similar to what Roth (2016) suggests—reasoning to be

social. The results confirmed that such background theory can help researchers to better understand students' IIR in a social perspective. The explicit use of background theories, as shown in this chapter, addresses a gap in statistics education research (see Nilsson et al. 2018). In this chapter, we illustrated that an inferentialist epistemology can constitute a meaningful supplement to ISI and IIR (as also suggested by Bakker and Derry 2011). Even though inferentialism also has its "dark spots" (as pointed out by Schindler et al. 2017), we think that it provides opportunities in particular for investigating IIR in order to inquire into students' reasoning as a social practice.

Acknowledgements This work was supported by the Swedish Research Council (Vetenskapsrådet) [2012-04811]. We furthermore want to thank the anonymous reviewers and especially the editors of this book for their constructive feedback and their efforts improving this article.

References

- Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal*, 32(1), 23–38.
- Ainley, J., Pratt, D., & Nardi, E. (2001). Normalising: Children's activity to construct meanings for trend. *Educational Studies in Mathematics*, 45(1–3), 131–146.
- 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*, 29(4), 455–470.
- Bakker, A., Derry, J., & Konold, C. (2006). Using technology to support diagrammatic reasoning about center and variation. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*. Voorburg, The Netherlands: International Association for Statistical Education and International Statistical Institute.
- Bakker, A., Kent, P., Noss, R., & Hoyles, C. (2009). Alternative representations of statistical measures in computer tools to promote communication between employees in automotive manufacturing. *Technology Innovations in Statistics Education*, 3(2).
- Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, 13(1–2), 5–26.
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45(1–3), 35–65.
- Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.
- Brandom, R. (2000). *Articulating reasons: An introduction to inferentialism*. Cambridge, MA: Harvard University Press.
- Brandom, R. (2001, July 12). Der Mensch, das normative Wesen. Über die Grundlagen unseres Sprechens. Eine Einführung. [The human, the normative being. About the foundations of our speech. An introduction.] *Die Zeit*. Retrieved from: https://www.zeit.de/2001/29/200129_brandom.xml.
- Burgess, T. A. (2006). A framework for examining teacher knowledge as used in action while teaching statistics. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*. Voorburg, The Netherlands: International Association for Statistical Education and International Statistical Institute.

- 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: Springer.
- 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.
- Gal, I., Rothschild, K., & Wagner, D. A. (1990). *Statistical concepts and statistical reasoning in school children: Convergence or divergence*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA, USA.
- 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.
- 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.
- Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction*, 14(1), 69–108.
- 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., & 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.
- Makar, K., & Rubin, A. (2018). Learning about statistical inference. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261–294). Cham: Springer.
- McClain, K., Cobb, P., & Gravemeijer, K. (2000). *Supporting students' ways of reasoning about data*. US Department of Education, Office of Educational Research and Improvement, Educational Resources Information Center.
- Meletioui-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.
- Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26(1), 20–39.
- Moritz, J. (2004). Reasoning about covariation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227–255). Dordrecht, NL: Springer.
- Moritz, J. B. (2000). Graphical representations of statistical associations by upper primary students. In J. Bana & A. Chapman (Eds.), *Mathematics Education Beyond 2000. Proceedings of the 23rd Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 440–447). Perth, WA: MERGA.
- Newen, A., & Schrenk, M. (2012). *Einführung in die Sprachphilosophie* [Introduction to the philosophy of language.]. WBG-Wissenschaftliche Buchgesellschaft.
- Nilsson, P., Schindler, M., & Bakker, A. (2018). The nature and use of theories in statistics education. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 359–386). Cham: Springer.
- Noorloos, R., Taylor, S., Bakker, A., & Derry, J. (2017). Inferentialism as an alternative to socioconstructivism in mathematics education. *Mathematics Education Research Journal*, 29(4), 437–453.
- Peregrin, J. (2009). Inferentialism and the compositionality of meaning. *International Review of Pragmatics*, 1(1), 154–181.
- Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education. Proceedings of the seventh international conference on teaching statistics, Salvador, Brazil*. Voorburg, The Netherlands: International Association for Statistical Education and International Statistical Institute.

- 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.
- 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. (1995). Young children's active and passive graphing. *Journal of Computer Assisted Learning*, 11(3), 157–169.
- Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 3–4.
- Rossmann, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5–19.
- Roth, W. M. (1996). Where is the context in contextual word problem?: Mathematical practices and products in grade 8 students' answers to story problems. *Cognition and Instruction*, 14(4), 487–527.
- Roth, W. M. (2016). The primacy of the social and sociogenesis. *Integrative Psychological and Behavioral Science*, 50(1), 122–141.
- Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa: International Association for Statistics Education.
- Schindler, M., Hußmann, S., Nilsson, P., & Bakker, A. (2017). Sixth-grade students' reasoning on the order relation of integers as influenced by prior experience: An inferentialist analysis. *Mathematics Education Research Journal*, 29(4), 471–492.
- Shaughnessy, M. (2006). Research on students' understanding of some big concepts in statistics. In G. Burrill (Ed.), *Thinking and reasoning with data and chance: 68th NCTM yearbook* (pp. 77–98). Reston, VA: National Council of Teachers of Mathematics.
- Watson, J. M. (2001). Longitudinal development of inferential reasoning by school students. *Educational Studies in Mathematics*, 47(3), 337–372.
- Wittgenstein, L. (1958). *Philosophical investigations*. Basil Blackwell.
- 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 8

Posing Comparative Statistical Investigative Questions



Pip Arnold and Maxine Pfannkuch

Abstract A “good” statistical investigative question is one that allows rich exploration of the data in hand, discovery, and thinking statistically. Two outcomes from four research cycles over a period of five years were: The development of criteria for what makes a good statistical investigative question and a detailed two-way hierarchical classification framework for comparative statistical investigative questions that are posed. With a focus on the last research cycle, responses from pre- and post-tests are explored, and the level of comparative statistical investigative questions that students posed is discussed.

Keywords Comparisons · SOLO taxonomy · Statistical enquiry cycle
Statistical investigative questions

8.1 Introduction

Arnold (2008) highlighted posing statistical questions as a problematic situation because of its role in assessment for qualifications in New Zealand and because teachers lacked knowledge in this area. The problem arose in the first of four research cycles where students in a test situation posed a statistical question, which was checked as satisfactory by the teacher. The students subsequently were unable to finish the test because their statistical question was not suitable for the given data. This raised the question “What makes a good statistical question?”, as the teacher and researcher together had marked the student posed questions correct. In an attempt to answer the question “What makes a good statistical question?” the literature was reviewed extensively and the conclusion drawn was that generally the literature gave

P. Arnold (✉)
Karekare Education, Auckland, New Zealand
e-mail: pip@karekareeducation.co.nz

M. Pfannkuch
Department of Statistics, The University of Auckland, Auckland, New Zealand
e-mail: m.pfannkuch@auckland.ac.nz

© Springer Nature Switzerland AG 2019
G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_8

mixed messages about what makes a good statistical question and the purpose of a statistical question. Indeed, Arnold (2013) concluded that the identified problem was actually about “What makes a good statistical *investigative* question?” Over four research cycles what makes a good statistical investigative question was explored, and the resultant criteria for what makes a good statistical investigative question were formed. These criteria informed the teaching experiment for research cycle four with a particular focus on comparative [statistical] investigative questions. Investigating comparative situations is a major focus in the New Zealand statistics curriculum at year 10 (ages 14–15) where this research took place. Hence, it is important for New Zealand teachers to know what makes a good statistical question at the school level, the components and concepts underpinning a good statistical question, and the learning in which students should be immersed to support the posing of good statistical questions. This chapter focuses on this fourth research cycle and explores the research question—*What level of comparative investigative questions are year 10 (ages 14–15) students posing?*

8.2 Literature Review

8.2.1 *Statistical Investigative Cycle*

The first dimension of the four-dimensional framework for statistical thinking in empirical enquiry (Wild and Pfannkuch 1999) is concerned with what one thinks about and the way in which one acts during a statistical investigation. Wild and Pfannkuch (1999) worked with the PPDAC (problem, plan, data, analysis, and conclusion) model (MacKay and Oldford 1994) of the statistical investigative cycle, and this is the model that underpins the work in this research:

- The **problem** stage deals with grasping a particular system’s dynamics and understanding and defining the problem.
- The **planning** stage involves deciding what to measure and how, how the sample will be taken, the design of the study, and how the data will be managed, including the recording and collecting of data. It also includes piloting the investigation and planning the analysis.
- The **data** stage is concerned with collecting, managing and cleaning the data.
- The **analysis** involves sorting the data, constructing tables and graphs as appropriate, exploring the data, looking for patterns, planned and unplanned analysis, and generating hypotheses.
- The final stage of the cycle involves interpreting, generating **conclusions**, new ideas and communicating findings.

In the statistical investigative cycle, questions and questioning arise in all areas. Questions are formally posed in both the problem and planning stages, in particular. Definitions and clarification of the purposes of these questions are now discussed.

8.2.2 Questions Within the Statistical Investigative Cycle

The initial motivating question for this research was: What makes a good statistical question? A number of studies were found where forming statistical questions were part of the researched process (e.g. Burgess 2007; Hancock et al. 1992; Lehrer and Romberg 1996; Pfannkuch and Horring 2005; Russell 2006) and a number of papers or books were located that reported an overview of the current status of statistics education, including forming statistical questions (e.g. Graham 2006; Konold and Higgins 2002; Whittin 2006). After reviewing existing literature and considering the statistical investigative cycle, the picture of what makes a good statistical question was still unclear. There were mixed messages about the purpose of statistical questions and whether they were used for an investigation or to collect data from people. From the literature (e.g. Burgess 2007; Russell 2006; Pfannkuch and Horring 2005) and from experience, it was concluded that within statistical investigations we can consider two types of questions: those that are formally posed and those that are spontaneously asked throughout the investigative process.

The theory proposed by Arnold (2013), therefore, is that there is *question posing* and *question asking*. Question posing results in a question being formally structured, whereas question asking is a continual spontaneous interrogative process. Question posing arises as a result of having a problem that needs to be addressed using a statistical investigation. Posed questions may be asked for investigative or survey purposes: *investigative* questions are those to be answered using data (the problem), while *survey* questions are those asked to get the data (the plan). Question asking also has two purposes, both of which involve an interrogation element: *interrogative* questions are those asked as checks within the PPDAC cycle, while *analysis* questions are those asked about the statistics, graphs and tables in order to develop a description of and an inference about what is noticed (the analysis).

As this research is focused on situations where students are working with secondary data, i.e. data that has already been collected and is given to them, Fig. 8.1 shows where questions fit within the statistical investigative cycle when students are given data. The purpose of Fig. 8.1 is to show how many different types of “statistical” questions are used within the PPDAC cycle, reinforcing how it could be very confusing for students if the questions are not defined and named according to their different purposes.

8.2.3 Posing Statistical Investigative Questions

In the big picture of statistical enquiry the investigative question is the statistical question or problem that needs answering or solving. In most instances the investigative question starts from an “inkling” and is developed into a precise question. The process of developing or creating the investigative question is iterative and requires considerable work to get it right (e.g., delMas 2004; Franklin et al. 2005;

| The PPDAC cycle in situations where the data are given to the students (secondary data). |
|---|
| <p>Original Investigators DATA/PLAN</p> <p>Data set is given.</p> <p><i>Interrogate</i> the background of the data.</p> <p>Understand the original <i>investigative</i> question(s) and data collection plan including how the data were collected and who the data were collected from. Find out the <i>survey</i> questions used. Identify the variables of interest and what they measure, and identify the population of interest.</p> <p><i>Interrogate</i> the data.</p> |
| <p>PROBLEM</p> <p>Motivating situation/question/idea of what “I” will investigate using this provided data.</p> <p>POSE a new <i>investigative</i> question.</p> <p><i>Interrogate</i> investigative question.</p> |
| <p>ANALYSIS</p> <p>Calculate statistics, draw graphs, build tables as appropriate.</p> <p>ASK <i>analysis</i> questions about the statistics, graphs and tables and describe what is noticed and what is inferred.</p> <p><i>Interrogate</i> analysis.</p> <p>Revisit the investigative question, updating if necessary (maybe pose new investigative questions that have become apparent after analysing the data).</p> |
| <p>CONCLUSION</p> <p>Write the “conclusion” answering the new investigative question(s).</p> <ul style="list-style-type: none"> • Use supporting evidence from the analysis. • Make inferences about the population. <p><i>Interrogate</i> conclusion.</p> <p>POSE further investigative questions as a result of “conclusion”.</p> |

Fig. 8.1 Questions within the statistical investigative cycle: secondary data (Arnold 2013, p. 22)

Hancock et al. 1992; Russell 2006; Wild and Pfankuch 1999). There is also a need when developing the investigative question to have “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” (Franklin and Garfield 2006, p. 350).

Posing investigative questions has been identified as a problem area for students, for example, the idea of asking questions of the data. Pfankuch and Horring (2005) noted that students lacked understanding of what a question is and the idea that one can pose a problem by asking questions of data: “Maybe students haven’t yet formed that understanding of what a question is—how you can ask a question in a set of data” (p. 208). Lehrer and Romberg (1996) found that students initially had problems with asking questions of data: “students believed that questions cannot be asked of data, only of people” (p. 80). Burgess (2007) noted that students found posing investigative questions a problem but did not specify the particular issue that arose. Other issues related to investigative questions include the need for teachers to model posing investigative questions, initially as seed or starter ideas (e.g. Lehrer and

Romberg 1996), but also to push students thinking about, for example, “typicalness” and data as an aggregate rather than individual cases (e.g. Konold and Higgins 2003).

In order to get precise investigative questions that can be interpreted and that yield useful information, an *interrogative* process, which involves asking questions of the investigative question, is necessary (e.g., Burgess 2007; Graham 2006; Konold and Higgins 2003). For example, Burgess (2007) acknowledged that some of the specialised content knowledge teachers needed for teaching statistics related to their ability to be able to decide if a question posed by their students was suitable, unsuitable, or whether changes could be made to make the question suitable. Graham (2006) provided five useful considerations for forming a good investigative question, which were different aspects of interrogating the investigative question. The considerations were whether the question was: “(1) actually a question, rather than simply an area for investigation...; (2) personally interesting to you...; (3) likely to draw on data that will be available within the time frame of the investigation...; (4) specific, so that it is answerable from data...; (5) measurable....” (p. 88). With this perspective, investigative questions are formulated through an interrogative process with regard to the considerations.

8.3 Methodology

Design based research (DBR) or design experiments were used. DBR has its foundations in design science (Brown 1992) and typically involves a planned intervention that develops ideas based on theoretically grounded innovations to inform practice while simultaneously conducting research on the intervention (Brown 1992; Cobb 2000). In particular, DBR focuses on the types of learning that differ from common or current practice and explores new and novel practices with the intent to change systems by being innovative (Bakker 2004; Bakker and van Eerde 2015; Schwartz et al. 2008).

A design experiment is a form of interventionist research that creates and evaluates novel conditions for learning. The desired outcomes include new possibilities for educational practice and new insights on the process of learning. Design experiments differ from most educational research because they do not study what exists; they study what could be. (Schwartz et al. 2008, p. 47)

The research, using DBR, started with an initial preparation and design phase, followed by a teaching experiment, then a retrospective analysis phase, which fed into another preparation and design phase, with the cycle repeated four times (e.g. Bakker 2004; Bakker and van Eerde 2015). A hypothetical learning trajectory (HLT) (Simon 1995) was used in the design of instructional materials. In the teaching experiment phase the teacher and researcher (as observer) together experienced the students’ learning and reasoning in the classroom. Each lesson was reflected on and informed the next lesson. During the teaching experiment phase, evidence was collected in the form of video-recordings of lessons, field notes, pre-and post-tests and interviews

of some students for the retrospective analysis, which occurred at two levels. An ongoing retrospective analysis informed subsequent planning and was motivated by what seemed best for the students (Cobb 2000). The retrospective analysis at the end of a teaching experiment was orientated by the HLT and conjectures both of which provided a basis for developing the instruction theory (Bakker 2004; Cobb 2000). The research process was iterative—design, test and redesign.

For the pre- and post-tests the retrospective analysis involved writing hierarchical descriptors based on the student data and criteria derived from the literature followed by the subsequent classification of student responses into categories. The categories evolved over four cycles and were based on the SOLO taxonomy (Biggs and Collis 1982). The SOLO taxonomy then provided the basis for quantification of the responses, which were then analysed quantitatively. Transcriptions of the video recordings were used to identify salient moments within the class lessons in order to provide evidence and illustrations of how students were scaffolded to interrogate and pose investigative questions.

Four research cycles were undertaken in 2007, 2008, 2009 and 2011. This chapter reports on the findings and outcomes from 2011, the last cycle. At the end of the first teaching cycle the problematic situation, what makes a good statistical question, was identified.

8.3.1 *Participants*

The first two teaching experiments were undertaken in a state, mid-socio economic, multicultural, suburban co-educational school with Teacher A, who in 2007 was in her fifth year of teaching. Her year 10 students (ages 14–15) in 2007 were average to below average in ability, while in 2008 the students were above average in ability. The last two teaching experiments were undertaken in a state, mid-socio economic, multicultural, inner-city girls' school with Teacher B, who in 2009 was in her ninth year of teaching. Her year 10 students in 2009 were average in ability. For the 2011 class focused on in this chapter, there were 29 students of above average ability involved in the research. The class had a mix of ethnicities including New Zealand European, Māori, Pasifika and Chinese.

8.4 Teaching Experiments

To situate the research question, the relevant elements of the four teaching experiments are given. These are the elements that: (1) contributed to the criteria; and (2) were relevant for comparative situations, the focus of the research question.

In every instance the teaching experiment is within the context of the statistics topic in a year 10 (ages 14–15) mathematics class. The main focus at this year level in New Zealand is on comparative situations. Generally the students would have about

4–5 weeks of the statistics topic across this one year of schooling. It is important for the reader to note that at the time of this research, the teaching of posing investigative questions in New Zealand was limited to teachers putting an investigative question on the board and then expecting the students to pose their own investigative questions with little or no formal teaching about how to pose investigative questions. For many students this would have been the first time they would have met comparative situations and especially the expectation to pose comparative investigative questions.

8.4.1 Teaching Experiment One

Posing investigative questions was identified at the end of the first teaching experiment as a problematic situation that was in need of further exploration (Arnold 2008). The hypothetical learning trajectory for posing investigative questions evolved over the teaching experiments. In the first teaching experiment, as questioning was not identified as a problematic area specifically, the teaching and learning sequence was created based on previous best practice, while focusing on using the statistical investigative cycle as envisioned in the new curriculum (Ministry of Education 2007). An initial linear hierarchical categorisation system was proposed for judging investigative questions (Arnold 2008) based on initial evidence in the students' post-tests [for a full account of student pre- and post-tests see Arnold (2008)].

8.4.2 Teaching Experiment Two

In the second teaching experiment the problematic situation, what makes a good investigative question, was initially addressed. During the teaching experiment the teacher focused on ensuring that the variable and the target population were clear in the question and that the question was asking about “some type of relationship or comparison” (Teacher A, 2008, lesson 2). In summarising questions within the statistical investigation cycle, three points were noted: (1) posing investigative questions requires students and teachers to have a clear idea of what the variable(s) are in which they are interested; (2) what they want to do (summarise, compare or relate); and (3) what the population of interest is. The planning involved deliberately teaching these criteria to the students and providing sufficient examples to allow them to practise with a number of different variables and populations. Teacher A deliberately discussed and highlighted the three criteria.

In addition to the initial lesson on posing investigative questions, the teacher decided to spend an additional lesson sorting, critiquing and improving investigative questions that had been posed by others. This involved the students first sorting the questions into the different types (summary, comparison and relationship) and then improving the investigative questions by making sure the investigative questions met the three criteria given by the teacher. In this lesson a number of points

| Criteria |
|--|
| <ol style="list-style-type: none"> 1. The variable(s) of interest is/are clear and available 2. The population of interest is clear 3. The intent is clear 4. The question can be answered with the data 5. The question is one that is worth investigating, that it is interesting, that there is a purpose 6. The question allows for analysis to be made of the whole group |

Fig. 8.2 Criteria for posing investigative questions

were mentioned by the teacher that have subsequently been linked to posing a good investigative question or understanding the question posed.

- The teacher mentioned several times during the lesson the need to consider whether the question was worth investigating. This links to Graham's (2006) second consideration (see Sect. 8.2.3).
- The actual variable that could be investigated was clarified; for example, they were not investigating foot size; they were investigating right foot length.
- The use of comparative words when posing comparison questions was explored to clarify the type of question; for example, using longer, taller or faster. Linked to this was the use of the appropriate comparing word (precise language) ; for example, use longer for right foot length, but not for right foot width (in this case they would use wider).

Between the second and third teaching experiments there was extensive dialogue between the researcher and colleagues at the university based largely on the retrospective analysis of student responses in the post-test from teaching experiment two. This dialogue addressed language and the preciseness of wording, in particular, the use of the article words *a* and *the* in investigative questions and the implications of these as to which group the question was about. Through this dialogue and through analysis of student responses, particularly poorly posed investigative questions, other ideas of suitable criteria for "What makes a good investigative question?" were generated. At this point six criteria were established (see Fig. 8.2) for what makes a good investigative question. These combine the three features the teacher used in the second teaching experiment, moderating questions from the first teaching experiment (Arnold 2008), and detailed analysis of the investigative questions that students posed in their pre- and post-tests.

The researcher then trialled some teaching ideas with a year nine (ages 13–14) class at another school to test how the criteria might be introduced to students. This was not recorded as it was not part of the research for which permission to video was granted, but it did provide an opportunity to trial some of the material before using it in the third teaching experiment.

8.4.3 *Teaching Experiment Three*

In the third teaching experiment criteria for what makes a good investigative question were used and the teaching focused on the underlying conceptual knowledge needed to understand the investigative question. The teaching and learning activities around posing investigative questions in the third teaching experiment built on the work from the second teaching experiment. The teacher (Teacher B, 2009) had been exposed to much deeper thinking about posing investigative questions prior to teaching the statistics unit. This had included workshops for all the teachers in the school on the material, and the teacher was a member of a linked Teaching and Learning Research Initiative project team (Pfannkuch et al. 2011).

Students posed investigative questions in class and then a selection of these was used in the following lesson. During the following lesson the teacher asked the students to sort questions that they had previously posed. The students identified which questions they thought were investigative questions and which ones were not. The students came to the conclusion that they did not like most of their questions. Through a teacher-led discussion the students generated ideas that aligned with the criteria for what makes a good investigative question. Students felt that the questions they had been given were not suitable as investigative questions: the question was not able to be answered because the variable was not one of the variables available in the given data set; there was not enough data to answer the question; and some questions were about an individual and not the whole group, which the students felt was unacceptable. Generating the criteria from student discussion and their findings was a deliberate strategy rather than the teacher just giving the criteria. An additional activity was used later in the topic where students critiqued questions that had been posed by others and improved on them based on the developed criteria.

As the teacher moved into new concepts, such as sampling, she started always with an investigative question, which was posed collectively as a class and checked against the criteria that had been established. In addition in later lessons on using samples to answer investigative questions about populations, care was taken to reinforce the actual population about which the students were posing and answering investigative questions. A fictitious school was invented and data cards for each “student” were created to help to develop the concept of population and sample. The “population”, Karekare College students, was constantly referred to, and this population was also physically shown as the data cards in a bag (see Fig. 8.3). This material representation of the population, coupled with the actual drawing of samples from the bag, was designed to reinforce the connection between sample and population and the investigative question.

In a wrap-up session the students again came back to the criteria about what makes a good investigative question, and, as well as posing investigative questions themselves, they had to critique questions posed by others. During this activity, an interesting observation was made by one of the students to another student in the group that was being observed: “Have you noticed that all the good ones are really long?” (2009 student, final lesson).



Fig. 8.3 Karekare College population bag with data cards (Arnold 2013, p. 152)

8.4.4 Teaching Experiment Four

In the final teaching experiment the teacher's (Teacher B, 2011) approach to posing investigative questions was different than in the previous experiments. She gave the students questions that had been posed by others before she required them to pose their own. The activity, where the students had to sort a number of investigative questions into groups, provided a catalyst to talk about what questions were good questions and what questions were not. From this discussion some of the criteria that had previously been established by the research were re-established by the students. That is, the students and teacher developed the criteria based on the class discussion about the questions they were sorting. Criteria that the students came up with included that the question needs to be about the overall distribution of the data, it must be interesting, and the variable and group need to be stated. Student reflection at the end of the lesson elicited a further criterion that had not been mentioned in class: that the type of question needed to be clear. At this point the teacher resisted the urge to "finish" the criteria (the students had identified five of the six criteria) and left the sixth criterion for when it naturally arose in the teaching and learning sequence.

Defining the context, i.e. the variable and the population, became a focus, and throughout the unit the teacher constantly asked the students to define the variable and the population for each situation. This was also linked to moving from questions about "these" students (the sample) to questions about the population. An example of the teacher helping the students define the variable is given in the excerpt below. It occurred in a lesson where students were exploring a situation where survey participants had ranked themselves as to how good they thought they were at a particular subject; for example, maths, reading, sport and the arts. The discussion was around exactly what the variable is, i.e. is it boys rating themselves higher than girls rate

themselves, or is it boys rating themselves as better when they compare themselves to girls?

Teacher: The question they were asked was how good do you think you are at maths. That was the question that they were asked. That was the survey question. ... How good do you think you are at maths? So remember we're comparing the boys and the girls. So when we're posing an investigative question we're looking at the first one, so those were the survey questions. The investigative question can someone give it to me, the first one?

Student: I wonder whether boys tend to think that they are better at maths than girls. Year 4–13 boys and girls.

Teacher: Year 4 to 13 New Zealand boys tend to think ...

Student: They are better at maths than girls.

Teacher: They are better at maths in this case than year 4 to 13 New Zealand girls.

Student: No not think they are. Because the boys wonder if they're better than the girls.

Teacher: Remember the question wasn't "Are you better than girls?", it's just how good you think you are so it's not rating against the other. But in the overall rating.

...

Teacher: What did we say up there? Boys rate themselves better at maths than girls. The boys aren't rating themselves compared to girls, it's just when they rate themselves, boys' ratings tend to be higher than girls' ratings. So the question could have been: "I wonder whether ratings for maths ability by year 4 to 13 New Zealand boys tend to be higher than ratings for maths ability by year 4 to 13 New Zealand girls."

In addition to the discussion regarding how to frame or describe the variable, the teacher was clearly differentiating between the two types of questions that are posed, i.e. survey questions and investigative questions. It is also worth noting her use of the phrase "tend to" for comparison questions. This phrasing had become part of the teacher's natural language she used in relation to comparison questions, a key element for a "good" comparison investigative question (see Fig. 8.4, and also links to criteria 6, Fig. 8.2).

The teacher persisted throughout the unit of work reinforcing the criteria for what makes a good investigative question, for example, getting the context sorted out by getting the students to correctly define the variable(s) and the population (criteria 1 and 2, Fig. 8.2), and making the questions about the population not the sample (criteria 2, Fig. 8.2). In addition she required them to make predictions of what they expected, particularly in the comparison situation, asking students all the time which group they thought would be bigger, taller or faster (links to criteria 3, 4 and 6, Fig. 8.2). The implication of these predictions was about how the comparison question was framed—for example, did they have boys taller than girls or girls taller than boys?—with the expectation being that the question was framed so that it aligned with what the students expected to be true. So if they thought boys were taller than

| Question category | Student question example | Commentary |
|---|--|--|
| A. Nonsense, not related or not a comparison question. | I wonder Auckland region and Wellington region have the same student in year 10? (2009 student, post-test) | This question is irrelevant and does not meet criteria 1, 3, 4 and 6 (Fig. 8.2). |
| | I wonder if the popliteal length relates to armspan. (2009 student, post-test) | This is a relationship question. |
| B. A question that is partially related to the data, but not answerable by the given data (either due to sample size issues or variable not in the data set). | I wonder if all the ambidextrous students are capable of kicking a ball with both left and right foot. (2009 student, pre-test) | Handedness was in the data set provided; however, there wasn't a question about ambidextrousness for "footedness". Therefore criteria 1 and 4 (Fig. 8.2) are not met. |
| | If Asian girls have a longer armspan than Indian boys. (2007 student, post-test) | In the 2007 post-test there were only two Indian boys and two Asian girls. Does not meet criteria 4 (Fig. 8.2). |
| C. A question that hints at comparison. | I wonder if more year 10 boys are physically fit than year 10 girls. (2011 student, post-test) | This question suggests comparison, though as it reads it is probably only comparing a couple of categories, therefore not meeting criteria 6 (Fig. 8.2). |
| | I wonder if ambidextrous hand writers can speak different languages. (2009 student, pre-test) | This question hints at comparing the number of languages spoken across handedness, and therefore not meeting criteria 3 and 4 (Fig. 8.2). |
| D. A question that has all of one group bigger/smaller than all of another group or compares an individual. | I wonder if all girls have longer hair than all boys. (2008 student, pre-test) | A good example of the type of thinking, and therefore the type of question, where students are thinking something is bigger and think all of one is bigger than all of the other. They have not yet grasped the idea of tendency or tending to be bigger/longer. |
| | I wonder if the average resting rate for a boy is lower than a girl? (2011 student, pre-test) | Comparing a boy with a girl, comparing individuals. |
| E. A question that compares categorical data. | I wonder if secondary students that live in southland region are fitter than secondary students from Auckland region. (2009 student, post-test) | In the data set given, the variables that might be used to answer this question were both categorical, region they live in, and fitness levels (unfit, a little fit, ...). |
| F. A question that compares a summary statistic. | I wonder if the typical right foot length for year 11 boys is greater than the typical right foot length for year 11 girls from the 2007 NZ CensusAtSchool database. (2008 student, post-test) | This question is comparing "the typical", which is interpreted as a summary statistic; for example, the median or the mode. |
| | I wonder if the average hair length of 16 year old girls is greater than the average hair length of 16 year old boys. (2008 student, pre-test) | This question is comparing the average, which could be median, mean or mode. |
| G. A question that assumes the idea of tendency. This includes questions that ask how much bigger or if there is a difference. | I wonder if secondary girl students have bigger wrist circumference than secondary boy students. (2009 student, post-test) | This question uses the phrase "have bigger" but, unlike the example in category D, they haven't indicated that they are thinking all girls bigger than all boys, so this style of question has been categorised as assuming tendency. |
| | I wonder if boys have longer popliteal lengths than girls. (2009 student, post-test) | A second example showing a different variable; commentary above relevant for this question. |
| H. A question that includes the idea of tendency; for example, question includes words or phrases such as on average, generally or tends. | I wonder if boys in year 10 tend to be taller than girls in year 10. (2009 student, post-test) | This question structure has one population tending to be taller/heavier OR have a longer/shorter [variable] than the other population. |
| | I wonder if on average right handers have longer hair than left handers. (2008 student, pre-test) | This is a similar structure to the first, but instead of using "tend", they have used "on average". |
| | I wonder if Yr 9-13 NZ boys have typically higher pulse rates compared to Yr 9-13 NZ girls. (2011 student, post-test) | This is a similar structure to the first also, but this time they have used "typically" to express the idea of tendency. |
| | I wonder if the popliteal length of Yr 9-13 NZ girls tend to be longer than Yr 9-13 NZ boys popliteal length (2011 student, post-test) | This question structure has the variable (of one of the groups) tending to be bigger/smaller than the variable (of the other group), a different structure to the previous three. |

Fig. 8.4 Comparison question examples (Arnold 2013, pp. 119–120)

girls, then the question was framed that way. A key realisation from this research was that students were conjecturing based on their general knowledge about which group would tend to have bigger values. The students were not explicitly aware they were making such a conjecture, but their posed questions strongly suggested that they were. The teacher was drawing on a new insight from the second and third teaching experiments.

8.5 Retrospective Analyses

Two findings came out of the retrospective analyses of student-posed investigative questions, these are: (1) criteria for what makes a good investigative question (Fig. 8.2, not the focus of the research question for this chapter); and (2) a detailed two-way classification matrix for comparative investigative questions that are posed (Sect. 8.5.4). The teaching experiments described above provide evidence of changes made to the teaching experiments as part of the ongoing retrospective analysis between cycles. This section describes in detail the retrospective analysis in relationship to the research question: What level of comparative investigative questions are year 10 (ages 14–15) students posing? The main sources of data were student pre- and post-test responses. In the pre- and post-tests the students were given a multivariate data set with 13 variables. Examples of discrete variables were: gender, year level at school and fitness level. Examples of continuous variables were: arm span, popliteal length (length from behind the knee to the floor, when the leg is bent at a right angle) and resting pulse rate. Students were asked to pose three comparison investigative questions.

8.5.1 *Classification of Comparison Investigative Questions*

A possible framework for comparison investigative questions was developed based on findings in the first teaching experiment (Arnold 2008). This initial framework considered questions that were not answerable with the data given and questions that were answerable, and the inclusion of the population signalled a higher level of question. This initial framework proved to be too simplistic, as it was found that the population descriptor required its own set of categories.

Comparison question categories were updated from the initial framework (Arnold 2008) following the second and third teaching experiments where student responses generated new categories. The categories were updated further following the fourth teaching experiment, as student responses signalled a need for further new categories. Figure 8.4 gives the final comparison question categories that were proposed for all year levels up to and including year 11. From year 12 onwards students have developed additional statistical knowledge which allows for more sophisticated investigative questions. This is not discussed in this chapter. Included in Fig. 8.4 are examples

for each of the different categories and commentary about the example to aid the reader. As signalled previously, the population descriptors (Criteria 2, Fig. 8.2) are not included. A good comparison investigative question needs to meet criteria 1, 3, 4, 5 and 6 from Fig. 8.2. Categories A–C are not comparison investigative questions, categories D and F are moving towards good comparison investigative questions, category E captures categorical data, which was considered inappropriate for the particular level and curriculum focus. Categories G and H are considered good comparison investigative questions with H being better than G.

8.5.2 Reflection on Final Framing of Comparison Investigative Questions

Two reflections on the final framing of the comparison investigative questions need to be mentioned. Firstly, the use of “tend to” to describe the idea of comparison, where one group “tends to be higher” than the other for a given variable, was signalled right from the start of the work on posing investigative questions (Pfannkuch et al. 2010). Researching students’ thinking about comparison situations (Pfannkuch 2006; Pfannkuch and Horring, 2005) had already identified “tend to” as being an important consideration in teaching thinking about the question framing for comparison situations.

Secondly, from the second teaching experiment to the third teaching experiment the framing of the question used in the pre- and post-tests moved from “I wonder if Year 11 NZ boys tend to have shorter **hair** than Year 11 NZ girls?” to “Does the **hair length** of Year 11 NZ boys tend to be shorter than the **hair length** of Year 11 NZ girls?” Both are acceptable as suitable comparison questions at this curriculum level, but the second question puts the variable (of the populations) clearly as the item that is being compared.

8.5.3 Population Descriptors

In the initial classification for comparison questions, the top category, H, was identified as being a “good” question and has the population included in the question. As student pre- and post-test responses were analysed from the second (2008) and third (2009) teaching experiments, it became clear almost immediately that the “super” category of population was not going to work. Students who had similar types of questions had a wide range of populations. For example, in the 2008 post-test 22 of the 24 students posed an investigative question about one group being taller than another group. Aside from the variation in the question format, 14 different population or group descriptors were used. The descriptors fell into three main categories:

(1) boys and girls (four variations); (2) various combinations of age groups (five variations); and (3) year 11 boys and girls (five variations).

Within the three broader categories there are multiple ways that students could phrase a descriptor based around whether they acknowledged that the broader population is New Zealand students and that the sample was taken from a particular CensusAtSchool database. It could be possible to make a fine graded scale for population descriptors, but pragmatism and what would be useful to teachers and students meant that fewer categories were better than more. Initially there seemed to be three clear categories: (1) Broad student population; for example, boys, girls, students (very general, could mean all boys and girls in the world); (2) Broad New Zealand student population; for example, New Zealand boys, New Zealand students (better than 1, doesn't consider the target subgroup of New Zealand boys and girls); (3) Actual New Zealand student population; for example, New Zealand year 10 students, New Zealand year 11 students, New Zealand secondary school girls.

However, as can be reasonably expected, student responses did not fall nicely into the three categories. Where, for example, did year 11 boys and year 11 girls fit? Clearly it is more specific than New Zealand students, but it doesn't specify New Zealand. An additional category was needed between broad New Zealand student population and actual New Zealand student population. Two other types of questions occurred that did not fit within these four categories. In the first type, students went broader than boys and girls but didn't use a specific population descriptor; for example, they asked about typical heights of males and females or of people. The second type of population descriptor that didn't fit into the four categories was when students specifically or inadvertently posed their investigative question about the sample. Examples of the second type of question are: "What are typical heights of **these** year 11 students?" and "What are typical heights for year 11 students sampled from the 2007 NZ C@S database?" Hence six population categories were considered as part of the overall question classification. These categories were confirmed through analysing the questions posed in the fourth teaching experiment. The final six population categories are:

1. Referring to the sample.
2. Broad population, not specifying students.
3. Broad student population; for example, boys, girls, students.
4. Broad New Zealand student population; for example, New Zealand boys, New Zealand students.
5. Any relevant student population that can be generalised about from the actual New Zealand student population used; for example, year 11 students, teenagers, secondary school girls.
6. Actual New Zealand student population; for example, New Zealand year 10 students, New Zealand year 11 students, New Zealand secondary school girls.

8.5.4 Two-Way Classification Framework (Matrix)

In order to classify a posed investigative question, the two categories, (1) question category and (2) population descriptor category, need to be considered as the student is working with both aspects (categories) at the same time. The combination of the two aspects gives rise to a two-way classification framework for comparison investigative questions. The framework is an 8 by 6 matrix (Fig. 8.5) made up of the eight question categories (rows, Fig. 8.4) and six population descriptor categories (columns, listed above).

The shaded portion of the matrix in Fig. 8.5 shows where the two aspects combine to give all the combinations to describe the investigative questions posed; for example, **H6** (in Fig. 8.5) is a comparative investigative question that includes the idea of tendency and has the actual New Zealand student population. The two-way classification framework developed during the retrospective analysis allowed for data to be gathered from each student to answer the research question—*What level of comparative investigative questions are year 10 (ages 14–15) students posing?*

8.6 Data Analysis

As described previously, students were asked to pose three comparison investigative questions in the pre- and post-test. These questions were each individually graded according to the comparison question category and the population descriptor category. For example, “*I wonder if the popliteal length of Yr 9–13 NZ girls tend to be longer than Yr 9–13 NZ boys popliteal length*” (student, post-test) was graded as **H6** because as a comparison question it *includes* the idea of tendency and it also has the actual New Zealand student population correct. On the other hand, “*I wonder if boys have longer popliteal lengths than girls*” (student, post-test) was graded as **G3** because it *assumes* the idea of tendency and has only specified a broad student population (boys and girls).

| | | Population categories | | | | | |
|--------------------------------|----|-----------------------|----|----|----|----|-----------|
| | | 1. | 2. | 3. | 4. | 5. | 6. |
| Comparison question categories | A. | A1 | A2 | A3 | A4 | A5 | A6 |
| | B. | B1 | B2 | B3 | B4 | B5 | B6 |
| | C. | C1 | C2 | C3 | C4 | C5 | C6 |
| | D. | D1 | D2 | D3 | D4 | D5 | D6 |
| | E. | E1 | E2 | E3 | E4 | E5 | E6 |
| | F. | F1 | F2 | F3 | F4 | F5 | F6 |
| | G. | G1 | G2 | G3 | G4 | G5 | G6 |
| | H. | H1 | H2 | H3 | H4 | H5 | H6 |

Fig. 8.5 Comparison investigative question matrix (Arnold 2013, p. 125)

The grading system gave 48 different possibilities when the question categories and the population descriptor categories were combined. In order to look at the difference from pre- to post-test, the 48 possibilities were simplified into six overall grades (see Fig. 8.6) using the SOLO taxonomy (Biggs and Collis 1982). The grades were based on the category of the question (A to H) and the population category (1 to 6). Pre-structural to extended abstract responses were scored from 1 to 5. Hence the above **H6** grade was scored as extended abstract or numerically as 5 while the **G3** grade was scored as multi-structural or numerically as 3.

A final pre-test and a final post-test score were determined for each student by finding the mean of their three SOLO grades for the three questions they posed. These final scores were analysed to look at the difference between pre- and post-test. Figure 8.7 shows the pre- and post-test questions posed by three different students chosen to give a range of responses, to demonstrate the grade given for each question and to show the subsequent SOLO score. For each question a student grade (a combination of comparison category and population) is given, for example E3, and their SOLO score for the question. This is summarised in the first column with their mean pre-test score, mean post-test score and the difference between the pre- and post-test means.

Student A moved from a combination of questions that were mostly non-comparison questions with a general student population to posing comparison investigative questions that include both the idea of tendency and the target population. Student A moved from pre-structural/uni-structural thinking to extended abstract thinking. Student B on the other hand was posing comparison questions, either comparing categorical variables or assuming the idea of tendency, but using a general student population. Student B moved to generally better population descriptors and also having more questions that included the idea of tendency, from multi-structural

| SOLO taxonomy level | Grade | Description of evidence |
|------------------------------|-------|---|
| No response or idiosyncratic | 0 | Questions that are not comparison questions, nonsense or not-related questions. Category A questions. |
| Pre-structural | 1 | Questions that are partially related to the data, but not answerable by the given data. Category B questions, any population. |
| Uni-structural | 2 | Questions that hint at comparison or have all of one group bigger/smaller than the other. Category C and D questions, any population. |
| Multi-structural | 3 | Questions that compare categorical data. Category E questions, any population. Relational or extended abstract categories (F, G and H) with population categories 1–4. |
| Relational | 4 | Questions that compare summary statistics or assume the idea of tendency, including the idea of difference. Population is “acceptable”. Category F and G questions with population categories 5 and 6. |
| Extended abstract | 5 | Questions that include the idea of tendency. Population is “acceptable”. Category H questions with population categories 5 and 6. |

Fig. 8.6 SOLO criteria for grading comparison investigative questions (Arnold 2013, p. 130)

| Student | Pre-test responses | Post-test responses |
|---------------------|--|--|
| Student A | I wonder whether the gender affects your fitness level (e.g. Are boys fitter than girls)? [E3, 3] | I wonder if the yr 9-13 NZ boys tend to have larger neck circumference than the yr 9-13 NZ girls. [H6, 5] |
| Mean pre-test: 1.7 | I wonder whether the armspan length is meant to be at a certain length whether you are a boy or girl? [C3, 1] | I wonder if the yr 9-13 NZ girls tend to have longer armspans than the Yr 9-13 NZ boys [H6, 5] |
| Mean post-test: 5 | I wonder whether the ring finger of the students are meant to be smaller than the index finger or not? [A3, 0] | I wonder if the yr 9-13 NZ boys tend to have larger popliteal lengths than Yr 9-13 NZ girls [H6, 5] |
| Difference: 3.3 | | |
| Student B | I wonder if the boy's wrist will be larger than the girls [G1, 3] | I wonder if yr 9-13 boys ringfinger at census [at] school tend to be longer than the yr 9-13 girls ring finger at census [at school] [H6, 5] |
| Mean pre-test: 2.7 | I wonder if more girls are less fit than boys [C3, 2] | I wonder if yr 9-13 girls tend to speak more languages than yr 9-13 boys at census [H6, 5] |
| Mean post-test: 4.7 | I wonder if girls are able to speak more languages compared to boys [G3, 3] | I wonder if yr 9-13 girls resting pulse is higher than yr 9-13 boys resting pulse at census [G5, 4] |
| Difference: 2 | | |
| Student C | I wonder what level of fitness most teenage boys are at [A3, 0] | I wonder if 2009 NZ C@S boys tend to have a longer armspan than 2009 NZ C@S girls [H4, 3] |
| Mean pre-test: 0 | I wonder what the average length of your index finger is for a teenage boy [A3, 0] | I wonder if 2009 NZ C@S girls tend to be more fit than 2009 NZ c@S boys [E4, 3] |
| Mean post-test: 3 | I wonder what the average pulserest is for teenage girls [A3, 0] | I wonder if 2009 NZ C@S boys tend to have a longer index finger length than 2009 NZ C@S girls [H4, 3] |
| Difference: 3 | | |

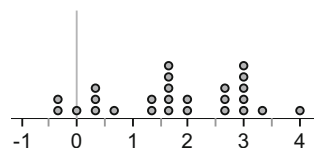
Fig. 8.7 Examples of student posed comparison investigative questions pre-test and post-test

thinking to extended abstract thinking. Student C initially was posing questions that were summary type questions, suggesting she did not understand what was meant by comparison questions. In the post-test student C was posing comparison questions and mostly ones that included the idea of tendency. This student still needed to work on her population descriptors because in all instances she was using the broad New Zealand student population rather than the target population.

8.7 Findings

From the class of 29 students, 26 students completed both the pre- and post-test. The findings are now discussed. Figure 8.8 shows the difference between students' pre-test mean score and their post-test mean score. A difference of two indicates that the student had a mean improvement of two points over their three comparison questions.

Fig. 8.8 Graph of difference between post-test mean score and pre-test mean score (Arnold 2013, p. 132)



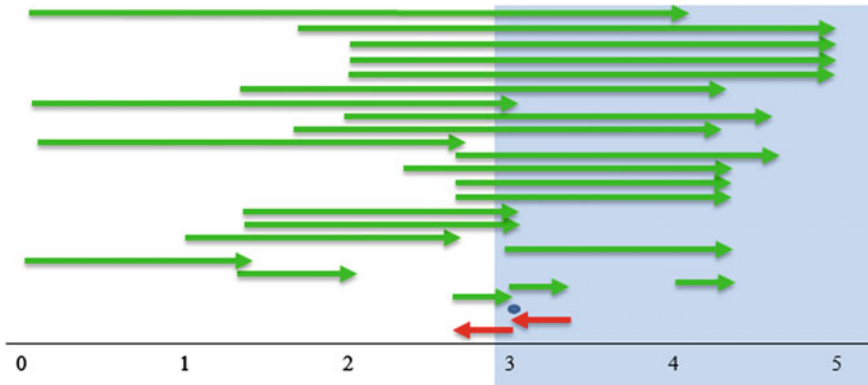


Fig. 8.9 Graph showing student movement from pre-test mean score to post-test mean score (red signals a negative movement—arrow pointing to the left, the circle signals no movement, green signals a positive movement—arrow pointing to the right) (color figure online)

Figure 8.9 shows the actual movement, from pre-test mean score to post-test mean score. The shaded grey area signals at least a multi-structural response on average. Of the 26 students 21 were posing at least at a multi-structural level on average in the post-test. The students in Fig. 8.9 are the same students as in Fig. 8.8. Working from left to right in Fig. 8.8 matches the students from bottom to top in Fig. 8.9.

Of the 26 students that sat both the pre- and post-tests, 23 improved their mean score (green/right pointing arrow in Fig. 8.9), one remained the same (circle in Fig. 8.9), and two lowered their mean score (red/left pointing arrow in Fig. 8.9). In the post-test, four students were working overall at extended abstract level, 10 at a relational level, seven at a multi-structural level; one at a pre-structural level, and four at a uni-structural level. The four uni-structural students all had a least one good question amongst their three, but were let down by a combination of the population category being low or one of the questions not being a comparison question. The pre-structural student asked questions that were about individuals (a boy, a girl) and also one non-comparison question. The students made significant improvement (p -value < 0.0001, paired t -test) in their mean scores from pre- to post-test question posing and on average increased their mean grade by 1.78 points (95% CI = [1.29, 2.28]).

Analysis of the different types of questions the 2011 students posed in their post-tests showed that for the comparison question categories (Fig. 8.10a) there were a higher proportion of questions in category H (45 questions out of 87) than any of the other categories for 2011. The population categories were also analysed across all the post-test questions (Fig. 8.10b).

No students used the sample (category 1) as the population, and only one student for one question used people generally as the population (category 2), and that was in just one question. The proportion of questions using acceptable populations (category 5 and 6) was 67.7%. While the good population descriptor did not always line up

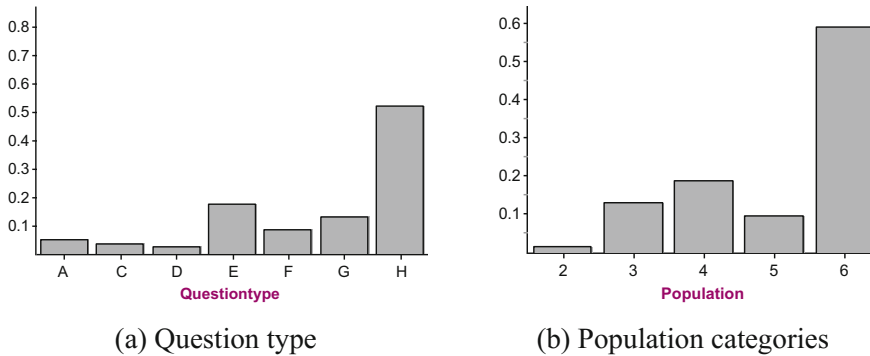


Fig. 8.10 Graphs of post-test analyses of comparison questions

with the good comparison question category, over half the question categories and two-thirds of the population descriptors were acceptable.

8.8 Discussion

The research question for this chapter was: What level of comparative investigative questions are year 10 (ages 14–15) students posing? The findings suggest that year 10 (ages 14–15) students are capable of posing comparative investigative questions that assume the idea of tendency (category H, Fig. 8.4) and have an acceptable population descriptor (Sect. 8.5.3); in other words they can pose “good” comparative investigative questions. For the 26 students who completed both the pre- and post-test, 54% were at least at this level. Most of the remaining students (27%) were posing comparative investigative questions, but their questions needed further refinement, mostly in terms of tidying up the population descriptor in the question.

There are considerations for statistics teaching and learning from the findings reported in this chapter. Firstly, the criteria (Fig. 8.2), the comparison question categories (Fig. 8.4) and the population descriptors (Sect. 8.5.3) provide structures to support teachers and students in improving their overall investigative question posing. If the quality of the question posed can be identified, for example, G4, then the improvements for the comparison question structure are given in Fig. 8.4 (G to H) and for the population descriptors (4 to 6). For students in particular, if they can become familiar with doing their own interrogation of their investigative questions against the criteria (Fig. 8.2), they will develop “thinking like a statistician” routines. Secondly, this chapter focuses only on investigative questions, the question that is asked of the data and implies a need for teachers to emphasise what a good investigative question is. Teachers also need to be discussing and highlighting the many other questions that are asked in statistical investigations, for example, survey questions, analysis questions, interrogative questions, inferential questions (Makar 2015). All

of these different question types make up the complex network of questions within the statistical investigation cycle. Thirdly, language used in investigative questions needs to be precise. Precise wording is critical (Biehler 1997; Pfannkuch et al. 2010) as “loose” or non-precise wording can cause confusion e.g. the use of *a* or *the* (see Sect. 8.4.2), and lead to poorly formed questions.

In addition, a number of statistical ideas and concepts should be developed concurrently. These include sample and population and the connection between the two, and ideas around tendency and typical. Finally, a statistical investigation is about more than just comparing or calculating simple measures; it is about students thinking distributionally, describing what they see in the sample(s) they have selected, and then making inferential statements about what may be happening back in the population(s) (Pfannkuch et al. 2010).

8.9 Implications

This research has identified gaps in the research knowledge base on posing statistical questions and consequently the big concepts underpinning the posing of good investigative questions that are needed for teaching and learning statistics at curriculum level 5 (ages 13–15) in New Zealand. This research into posing investigative questions has already had a huge impact in New Zealand classrooms and beyond year 10, curriculum level 5 (ages 13–15). Posing investigative questions is a key aspect of many of the statistics achievement standards in the national assessments, and the term investigative question is now widely used. Criteria for what makes a good investigative question, along with summary and comparison question categories, are available online as a support for teachers.

Implications for teachers include having the opportunity to experience the teaching and learning material in order to support their understandings of the research findings. Ideally this needs to happen before they take the material into their classrooms to use with their students. The sharing of the findings can support the teachers in the same way as it is hoped they will help their students. In addition supporting teachers to understand the different purposes of questions in statistics or how the use of precise language is important and making them aware of the potential confusions in language use for students would also be essential components of any work with teachers.

There is an urgent need to upskill teachers in their knowledge of the conceptual foundations required for posing good statistical questions. Many mathematics and statistics teachers are mathematics—not statistics—trained, or trained years ago. Either way, the statistics of today is not the statistics of their schooling or university days. It requires new knowledge and new ways of thinking. It also requires new ways of teaching, from a focus on the skills and calculations of the old statistics curriculum to a focus on the statistical reasoning and thinking that is inherent in the new New Zealand statistics curriculum as well as in the curricular guidelines for many other countries, such as GAISE in the United States (Franklin et al. 2005).

This chapter has addressed the problematic situation around what makes a good investigative question, including the underpinning concepts that are needed to support the teaching and learning. Suggested further research could include a focus on interrogating the statistical investigative cycle to find out what aspects should be a focus for students ages 13–15 or at other ages. Another suggestion for future research could be to explore students asking analysis questions. For example, what thinking prompts do students need to have when they are starting to analyse their data? Also because this research focused on comparison investigative questions and related research has explored summary investigative questions, three further areas of research could be posing relationship, time-series, and two-way table (two categorical variables) investigative questions.

References

- Arnold, P. (2008, July). *What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation*. Paper presented at the 11th International Congress on Mathematical Education (ICME-11), Monterrey, Mexico. <http://tsg.icme11.org/document/get/481>.
- 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>.
- Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: Freudenthal Institute.
- Bakker, A., & van Eerde, D. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbals, 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.
- 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. Proceedings of the International Association for Statistical Education Round Table Conference*, July, 1996. Granada, Spain (pp. 169–190). Voorburg, The Netherlands: International Statistical Institute.
- Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York, NY: Academic Press.
- Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2.
- Burgess, T. (2007). *Investigating the nature of teacher knowledge needed and used in teaching statistics* (Doctoral thesis). Retrieved from <http://www.stat.auckland.ac.nz/~iase/publications/dissertations/07.Burgess.Dissertation.pdf>.
- Cobb, P. (2000). The importance of a situated view of learning to the design of research and instruction. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 45–82). Westport, CT: Ablex.
- 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.
- Franklin, C., & Garfield, J. (2006). The GAISE project. Developing statistics education guidelines for grades Pre-K–12 and college courses. In G. Burrill & P. Elliot (Eds.), *Thinking and reasoning*

- with data and chance: *Sixty-eighth yearbook* (pp. 345–375). Reston, VA: National Council of Teachers of Mathematics.
- 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.
- Graham, A. (2006). *Developing thinking in statistics*. London, England: Paul Chapman.
- Hancock, C., Kaput, J., & Goldsmith, L. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 2(3), 337–364. https://doi.org/10.1207/s15326985ep2703_5.
- Konold, C., & Higgins, T. (2002). Highlights of related research. In S. J. Russell, D. Shifter, & V. Bastable (Eds.), *Developing mathematical ideas: Collecting, representing, and analyzing data* (pp. 165–201). Parsippany, NJ: Dale Seymour.
- 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.
- Lehrer, R., & Romberg, T. (1996). Exploring children’s data modeling. *Cognition and Instruction*, 14(1), 69–108.
- MacKay, R., & Oldford, W. (1994). *Stat 231 course notes fall 1994*. Waterloo, Canada: University of Waterloo.
- Makar, K. (2015). *Informal inferential reasoning*. Keynote presentation to the Combined Hui, Auckland, New Zealand.
- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, New Zealand: Learning Media.
- Pfannkuch, M. (2006). Comparing boxplot distributions: A teacher’s reasoning. *Statistics Education Research Journal*, 5(2), 27–45.
- Pfannkuch, M., Arnold, P., & Wild, C. J. (2011). *Statistics: It’s reasoning not calculating* (Summary research report on Building students’ inferential reasoning: Levels 5 and 6). Retrieved from <http://www.tlri.org.nz/tlri-research/research-completed/school-sector/building-students-inferential-reasoning-statistics>.
- Pfannkuch, M., & Horring, J. (2005). Developing statistical thinking in a secondary school: A collaborative curriculum development. In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education 2004 round table* (pp. 204–218). Voorburg, The Netherlands: International Statistical Institute.
- Pfannkuch, M., Regan, M., Wild, C. J., & Horton, N. (2010). Telling data stories: Essential dialogues for comparative reasoning. *Journal of Statistics Education*, 18(1). <http://www.amstat.org/publications/jse/v18n1/pfannkuch.pdf>.
- Russell, S. J. (2006). What does it mean that “5 has a lot”? From the world to data and back. In G. Burrill & P. Elliot (Eds.), *Thinking and reasoning with data and chance: Sixty-eighth Yearbook* (pp. 17–29). Reston, VA: National Council of Teachers of Mathematics.
- Schwartz, D., Chang, J., & Martin, L. (2008). Instrumentation and innovation in design experiments. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 45–67). New York, NY: Routledge.
- Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, 26(2), 114–145.
- Whitin, D. (2006). Learning to talk back to a statistic. In G. Burrill & P. Elliot (Eds.), *Thinking and reasoning with data and chance: Sixty-eighth yearbook* (pp. 31–39). Reston, VA: National Council of Teachers of Mathematics.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–265. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>.

Part III
Teachers' Knowledge
(Preservice and Inservice)

Chapter 9

Pre-service Teachers and Informal Statistical Inference: Exploring Their Reasoning During a Growing Samples Activity



Arjen de Vetten, Judith Schoonenboom, Ronald Keijzer and Bert van Oers

Abstract Researchers have recently started focusing on the development of informal statistical inference (ISI) skills by primary school students. However, primary school teachers generally lack knowledge of ISI. In the literature, the growing samples heuristic is proposed as a way to learn to reason about ISI. The aim of this study was to explore pre-service teachers' reasoning processes about ISI when they are engaged in a growing samples activity. Three classes of first-year pre-service teachers were asked to generalize to a population and to predict the graph of a larger sample during three rounds with increasing sample sizes. The content analysis revealed that most pre-service teachers described only the data and showed limited understanding of how a sample can represent the population.

Keywords Informal inferential reasoning · Informal statistical inference
Initial teacher education · Primary education · Samples and sampling
Statistics education

A. de Vetten (✉) · B. van Oers
Section of Educational Sciences, Vrije Universiteit Amsterdam, Amsterdam,
The Netherlands
e-mail: a.j.de.vetten@fsw.leidenuniv.nl

B. van Oers
e-mail: bert.van.oers@vu.nl

J. Schoonenboom
Department of Education, University of Vienna, Vienna, Austria
e-mail: judith.schoonenboom@univie.ac.at

R. Keijzer
Academy for Teacher Education, University of Applied Sciences iPabo, Amsterdam,
The Netherlands
e-mail: r.keijzer@ipabo.nl

© Springer Nature Switzerland AG 2019
G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_9

9.1 Introduction

In today's society, the ability to reason inferentially is increasingly important (Liu and Grusky 2013). One form of inferential reasoning is informal statistical inference (ISI), which is defined as “a generalized conclusion expressed with uncertainty and evidenced by, yet extending beyond, available data” (Ben-Zvi et al. 2015, p. 293) without the use of formal statistical tests based on probability theory (Harradine et al. 2011). In recent years, statistics education researchers have focused on how primary school students can be introduced to ISI. Scholars have hypothesized that if children are familiarized with the concept in primary school, they will understand the processes involved in ISI reasoning and in statistical reasoning in general (Bakker and Derry 2011; Makar et al. 2011). Evidence suggests that meaningful learning environments can render ISI accessible to primary school students (Ben-Zvi et al. 2015; Meletiyou-Mavrotheris and Papanistodemou 2015).

If children are to be introduced to ISI in primary school, future teachers need to be well prepared to provide this introduction (Batanero and Díaz 2010). They must have appropriate knowledge of the field that extends beyond the students' knowledge (Burgess 2009). It has been shown, however, that pre-service teachers' knowledge of ISI is generally weak (Batanero and Díaz 2010; De Vetten et al. 2018). This points out the need to improve the ISI content knowledge of pre-service teachers.

Current research provides only scant evidence for how to support the development of pre-service teachers' knowledge of ISI (Ben-Zvi et al. 2015). In some statistics education literature, the growing samples heuristic is recommended to stimulate ISI reasoning (Joan Garfield et al. 2015). The idea of this heuristic is that samples of increasing size are used to make inferential statements about a larger sample or population. Using this heuristic to informally and coherently construct and discuss ISI has typically not been investigated in the context of teacher education. Therefore, we implemented the growing samples heuristic in three classes of first-year pre-service teachers and explored their ISI reasoning when engaged in an activity that applies this heuristic.

9.2 Theoretical Background

9.2.1 Teachers' Knowledge of ISI

Teachers need to possess thorough knowledge of the content they teach (Hill et al. 2008) that extends beyond what their students actually learn (Ball et al. 2008), since the former's content knowledge impacts the latter's learning achievements (Rivkin et al. 2005) and facilitates the development of pedagogical content knowledge (Ball et al. 2008; Shulman 1986). It has been shown that these relationships also hold for ISI (Burgess 2009; Leavy 2010).

To conceptualize the required knowledge of ISI for pre-service teachers, we used the Makar and Rubin (2009) ISI framework. The three components of this framework are broad to include various types of students (Makar and Rubin 2014). For this study among pre-service teachers, we conceptualized the components in the following way:

1. “Data as evidence”: The inference needs to be based on the data and not on tradition, personal beliefs or experience. To base an inference on the sample data, the data need to be analyzed descriptively, for example, by calculating the mean (Zieffler et al. 2008). The resulting descriptive statistic then functions as an evidence-based argument within ISI (Ben-Zvi 2006).
2. “Generalization beyond the data”: The inference goes beyond a description of the sample data to make a claim about a situation beyond the sample data.
3. “Probabilistic language”: The inference includes a discussion of the sample characteristics, such as the sample size and sampling method, and what these characteristics imply about the representativeness of the sample and the certainty of the inference. Moreover, the inference requires understanding whether a sample is properly selected, the sample-to-sample variability is low, and this sample is representative of the population and can be used for an inference.

One of the studies that have investigated (pre-service) primary school teachers’ content knowledge is De Vetten et al. (2018). In a large-scale questionnaire study, they found that about half of the pre-service teachers agreed that data can be used as reliable evidence for a generalization. The authors also showed that the respondents were able to discern that probabilistic generalizations are possible, while deterministic generalizations are not. The evidence for the Probabilistic language component suggests that many pre-service teachers have a limited understanding of sampling methods, sample size, representativeness and sampling variability (De Vetten et al. 2018; Meletiou-Mavrotheris et al. 2014; Mooney et al. 2014; Watson 2001). With respect to the knowledge of descriptive statistics more generally, (pre-service) teachers’ knowledge has been shown to be typically superficial (Batanero and Díaz 2010; Garfield and Ben-Zvi 2007; Jacobbe and Carvalho 2011). More specifically, pre-service teachers tend to focus on measures of central tendency at the expense of measures of dispersion (Canada and Ciancetta 2007); while the group’s understanding of the mean, median and mode is mostly procedural (Groth and Bergner 2006; Jacobbe and Carvalho 2011). De Vetten et al. (2018) asked respondents to evaluate which descriptive statistics were well suited as arguments within ISI. The respondents acknowledged that ISI can be based on global descriptive statistics, but they did not recognize that ISI based on local aspects of the sample distribution is not correct. These studies indicate that there is a need to improve pre-service teachers’ ISI content knowledge.

9.2.2 Using the Growing Samples Heuristic to Support the Development of ISI

Research on how to support pre-service teachers' development of ISI content knowledge is almost nonexistent (Ben-Zvi et al. 2015). Leavy (2006) intervention study examined pre-service primary school teachers' distributional reasoning when engaged in experimental investigations. She found that the pre-service teachers in the sample tended to compute measures of centrality only rather than explore datasets, for example, using graphical representations. Moreover, the teachers often neglected the role of variation in comparing distributions. However, the participants became more attentive to variation and looked more at aggregate features of the distributions. Although the tasks used were inferential, the analysis focused on distributional reasoning only. Leavy (2010) showed that final-year pre-service teachers do not reflect on the meaning of the graphs and calculations they perform. The activity at the start of the intervention involved making inferences and discussing sampling issues, but the author did not analyze the activity in depth. These studies revealed that the pre-service teachers in her sample tended to restrict their attention to descriptive statistics, rather than how these descriptive statistics can be used in ISI.

In the context of statistics education generally, the growing samples heuristic has been suggested as a promising approach to support the development of ISI reasoning (Joan Garfield and Ben-Zvi 2008; Garfield et al. 2015). The idea of this heuristic is that samples of increasing size are used to make inferential statements about a larger sample or population. Ben-Zvi et al. (2012) showed that the heuristic helps middle-grade students not only describe samples but also draw conclusions beyond the data. Moreover, these students' reasoning about uncertainty developed from either certainty only or uncertainty only to more sophisticated reasoning in which probability language was used. Bakker (2004) found that when middle-grade students use this heuristic, they develop coherent reasoning about key distributional aspects of samples, such as center, spread and density. We hypothesize that the growing samples heuristic can aid the use of data as evidence because this heuristic draws students' attention repeatedly to the data. Research suggests it is not self-evident that students see sample data as evidence from which to make generalizations and predictions (Ben-Zvi et al. 2007; Makar and Rubin 2009). The heuristic could also help students to understand sample-to-sample variability because as the sample size increases, the shape of the distribution stabilizes and more likely resembles the population distribution (Garfield and Ben-Zvi 2008; Konold and Pollatsek 2002). In our view, the heuristic may be well suited for teacher education, because the relative simplicity of the heuristic allows pre-service teachers to translate the growing samples activities to their own teaching practice in primary schools.

9.2.3 Research Aim and Question

Until now, little, if anything, is known about how the growing samples heuristic supports pre-service teachers' development of reasoning about ISI. We hypothesize that pre-service teachers may reason differently from middle school students. On the one hand, we hypothesize that pre-service teachers are better suited to reason about ISI because they have more (procedural) statistical knowledge, which they can use in reasoning about ISI. Moreover, given their older age, they may be more able to reason about an abstract population. On the other hand, their future role as teachers might hinder them in drawing inferences as they might have a class of children in mind as their natural population of interest (Schön 1983). Therefore, pre-service teachers could relate sample results to a class instead of to an abstract population.

The aim of this exploratory study was to investigate the reasoning about ISI of 35 pre-service primary school teachers divided over three classes when they were engaged in a growing samples activity (Fig. 9.1). The research question is: What reasoning about informal statistical inference do first-year pre-service primary school teachers display when they are engaged in a growing samples activity, and what is the quality of their reasoning?

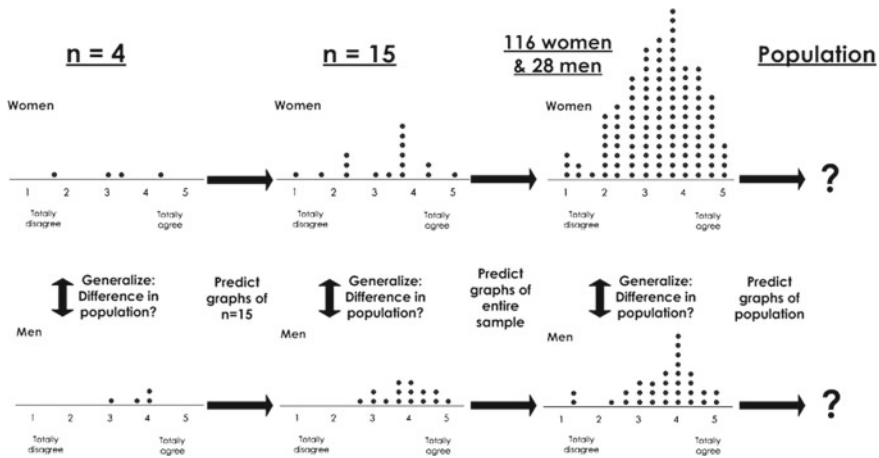


Fig. 9.1 Growing samples activity used in the current study

9.3 Methods

9.3.1 Intervention

The three components of ISI provided the framework for the pre-service teachers' ISI learning objectives. We formulated 10 learning objectives (Table 9.3, the last column), which informed the design of the growing samples activity (Fig. 9.1). The activity was inspired by the activities used by Bakker (2004) and Ben-Zvi et al. (2012) and consisted of three rounds. In each round, the participants answered the question, "Is the attitude toward mathematics of first-year male pre-service teachers in general more positive than the attitude toward mathematics of first-year female pre-service teachers?" Before the participants analyzed the data, they discussed how the data were collected and how the data could be used to answer this question, the "talking through the data creation process" (Cobb and Tzou 2009). This process was used to support the participants' confidence in the validity of their conclusions. The teacher educator stressed that the question pertained to the population of all Dutch first-year pre-service teachers and explained that the data came from a research project conducted among pre-service teachers at their teacher college the previous year. The data showed the averages of three 5-point Likert items. Therefore, the data could take on values between one and five with increments of one third. Next, the participants were provided with graphs of samples of increasing size. During each round, the participants answered the question about the difference in the population and predicted the shape of the graphs of the next round. The sheet of paper on which the participants filled in their answers to these questions also showed the graphs of the particular round. During the first round, the samples consisted of four men and four women, during the second round 15 men and 15 women, and during the third round 28 men and 116 women, which was the size of the original dataset. The sample size of four was meant to elicit responses of high or even complete uncertainty. The samples sizes of the second and third rounds were chosen to investigate whether the certainty of the participants' responses would increase. Round 3 also provided the opportunity to discuss ways to compare samples of unequal sizes. After each round, the answers were discussed in a class discussion. This discussion had similar patterns for each round in each class: The teacher educator asked for an answer to the question, asked on what grounds this answer was reached and probed for the certainty of the conclusion. Next, the prediction for the larger samples was discussed. During the last round, the comparison of the samples of unequal sizes was discussed, and the arguments used during the entire activity were summed up. Some parts of the discussion were more extensive than others, depending on the input from the pre-service teachers. The activity lasted for 80 min. All three rounds were held on one day. The participants worked in groups of two or three (Class A: 7 groups; Class B: 5 groups; Class C: 4 groups). In Class C, the first round was skipped because during the third round in classes A and B motivation seemed to decline.

For each learning objective, we formulated a conceptual mechanism that explained how the activity was hypothesized to scaffold the participants' reasoning to attain the

learning objectives (see Table 9.3, the conceptual mechanism column). We hypothesized that the repetition of the question to generalize and predict would invite the participants to use the data to draw a conclusion (the Data as evidence component). When the sample sizes increased, the averages and the global shape stabilized. We expected that the participants would notice that this and would use them as reliable signals for generalization and prediction. The repetition of the questions would also draw the participants' attention to the inferential nature of the question (component Generalization). Furthermore, we expected that presenting samples of different sizes and shapes would draw the participants' attention to differences in the sample distributions and would influence the participants to realize that other sample distributions could have resulted as well. They would, in turn, take the uncertainty of their conclusions into account. Finally, comparing how the sample data were spread about the center of the data of the various samples would encourage the participants to take uncertainty into account (the Probabilistic language component).

9.3.2 Participants

Three classes (A, B and C) for a total 35 first-year pre-service primary education teachers participated in this study. This was a convenience sample, as the first author also taught their course on mathematics education. They attended a small teacher college in a large city in the Netherlands. In the Netherlands, initial teacher education starts immediately after secondary school and leads to a bachelor's degree. For these students, mathematics teaching is usually not their main motive for becoming teachers. The mean age of the participants was 19.47 years (SD: 1.54), three were male, 20 had a background in secondary vocational education (students attending this type of course are typically between 16 and 20 years old), 13 came from senior general secondary education, and the educational background of the remaining two was either something else entirely or unknown. Table 9.1 shows the educational backgrounds for each class. Whereas descriptive statistics, probability theory and some inferential statistics are part of the mathematics curriculum of senior general education, these topics are generally not taught in secondary vocational education.

9.3.3 Data Collection and Data Analysis

Data collection consisted of the participants' answer sheets and sound recordings of the class discussions.

Content analysis in Atlas.ti was used to analyze the data. A coding scheme was developed based on the learning objectives. All answer sheets (both text and graphs) and class discussions were coded by assigning one or more codes related to the learning objectives to the data. In an iterative process, the first author and an external coder coded and discussed the coding scheme instructions until the instructions

Table 9.1 Educational background per class of pre-service teachers

| Educational background | Class A | Class B | Class C | Total |
|--------------------------------|---------|---------|---------|-------|
| Secondary vocational education | 7 | 9 | 4 | 20 |
| Senior general education | 7 | 2 | 4 | 13 |
| Something else/unknown | 1 | 0 | 1 | 2 |
| Total | 15 | 11 | 9 | 35 |

were deemed clear enough to be applied by a second coder. First, for each round all contributions of each group or participant were put in a table, organized by learning objective. Second, the group or individual contributions were aggregated for each class, organized by learning objective. Third, to measure the quality of the participants' reasoning about ISI, for each class, the contributions per learning objective were compared to the hypothesized conceptual mechanism. This comparison was conducted for each round and separately for the answer sheets and the class discussions and for each class. Per round, an indicator (—, —, 0, + or + +) was assigned to each learning objective, indicating to what extent the actual reasoning of a class was in line with the hypothesized reasoning (see Table 9.2). The indicators served as quality indicators of the classes' reasoning about ISI. Next, the indicators per round were combined into one indicator per learning objective for the three rounds together, separately for the answer sheets and the class discussions. Finally, the separate indicators for the answer sheets and the class discussions were combined into one indicator per learning objective. These combined indicators were used to compare the reasoning between the three classes. The assignment of the indicators was discussed with an external researcher until consensus was reached (Table 9.3).

9.4 Results

9.4.1 Answer Sheets

Table 9.4 shows the results of the answer sheets, aggregated over the three classes to give a comprehensive picture of the results. Aggregation was possible because the differences between the three classes' answer sheets were small. The left part of Table 9.4 shows how the participants answered the question to draw a conclusion about the difference between men and women in the population; the right part shows how the participants predicted the distribution of a larger sample or the population. In general, most participants used the data as evidence for their conclusions. In 36 of

Table 9.2 Meaning of indicators assigned to indicate the quality of classes' reasoning about ISI

| Indicator | Meaning |
|-----------|--|
| ++ | The reasoning of all pre-service teachers is in line with the learning objective. |
| + | The reasoning of about half (Class A: 3 of 7 groups; B: 2 or 3 of 5 groups; C: 2 of 4 groups) of the pre-service teachers is in line with the learning objectives, whereas there is insufficient evidence about the other pre-service teachers; or: 1 or 2 (Class A) or 1 (B and C) groups show misconceptions regarding the learning objective, while the other groups reason in line with the learning objective. |
| 0 | About half (A: 3 groups; B: 2 or 3 groups; C: 2 groups) of the pre-service teachers show reasoning that is and that is not in line with the learning objective; or: 1 or 2 (A) or 1 (B and C) groups of pre-service teachers reason in line with the learning objectives, whereas there is insufficient evidence about the groups. |
| - | The reasoning of about half (A: 3 groups; B: 2 or 3 groups; C: 2 groups) of the pre-service teachers is not in line with the learning objectives, whereas there is insufficient evidence about the other pre-service teachers; or: 1 or 2 (A) or 1 (B and C) group shows reasoning that is in line with the learning objectives, while the reasoning of the other groups is in not line with the learning objective. |
| -- | The reasoning of all pre-service teachers is not in line with the learning objective, evidenced from misconceptions, or from no attention for the learning objective, where attention was expected. |

Table 9.3 Conceptual mechanism and learning objectives for the growing samples activity

| Component | Aspect | Aspect of the growing samples activity | Conceptual mechanism | Learning objective |
|------------------|------------------|---|--|--|
| Data as evidence | Data as evidence | PTs ^a are repeatedly asked to draw conclusions. | The repetition of the question creates awareness that empirical data can be used to base conclusions on. | PTs base their conclusions on evidence from the data, not on context-based claims, such as preconceived ideas. |
| | Center | PTs are asked to generalize to a population and to predict the shape of a larger sample or of the population. | To answer the questions to generalize or to predict, PTs use conceptual tools as data-based arguments. These tools are statistical measures of aspects of the sample distribution. One measure is the mean, which is a stable feature of the distribution. Another is the mode, which is not stable for smaller samples but can function as an indicator for the majority. | PTs compare samples by using measures of centrality that are familiar to them, such as the mean (see Sampling variability). They either estimate or calculate the mean. To predict, they use approximately the same mean, mode and global shape as in the smaller sample. |
| | Spread | PTs are asked to generalize and to predict based on samples with varying spreads. | When PTs are confronted with only one sample, they may not consider the spread. Presenting samples with different spreads creates awareness about the spread of the sample distributions. | PTs employ different statistical measures for the spread, such as the range, or informal statistical expressions for the spread, such as more spread out. These spread measures are used in the PTs' generalizations (see Heterogeneity) but also as a signal for the spread of the larger sample sizes. |

(continued)

Table 9.3 (continued)

| Component | Aspect | Aspect of the growing samples activity | Conceptual mechanism | Learning objective |
|--------------------------------|----------------|---|--|--|
| Generalization beyond the data | Distribution | PTs are asked to predict the shape of a larger sample or of the population. | The question to predict the shape of a larger sample helps PTs to focus on the general shape and on aggregate features of the distribution. When the samples sizes increase, PTs will notice that the shape of the distributions stabilizes (i.e. does not change much). | Based on visual inspection of the graphs, PTs describe the general shape of the sample distributions in informal statistical expressions, such as majority, minority, few, most are left, and positive. PTs reason about the general shape and about the aggregate features of the distribution. In the predictions, PTs sketch graphs with approximately the same shape as in the previous round. |
| | Generalization | PTs are repeatedly asked to generalize to the population. | If the question to generalize is asked only once, PTs may describe only the sample at hand. Repetition will alert PTs to the inferential aspect of the question. | PTs make probabilistic generalizations. |
| Probabilistic language | Prediction | PTs are asked to predict the shape of a larger sample or to the population. | The question itself obliges PTs to generalize beyond the data, since PTs are asked to expand the graph with additional data. | PTs generalize to a larger sample or to the population, copying the general shape of the smaller sample, with approximately the same mean and mode (see Center). |
| | Sample size | PTs are asked to compare samples of different sizes. | When the sample sizes increases, PTs notice that the shape of the distribution does not change much from a certain sample size onward. | PTs conclude that larger sample sizes provide more certainty about their inferences, because some sample features appear to be stable. |

(continued)

Table 9.3 (continued)

| Component | Aspect | Aspect of the growing samples activity | Conceptual mechanism | Learning objective |
|----------------------|--|--|---|--|
| Heterogeneity | PTs are provided with samples with different spreads. | PTs are provided with samples with different spreads. | PTs compare samples where the data is spread differently about the center of the data and understand that sample distributions with large spreads provide less information about whether two population means differ. | PTs conclude that the more the sample data are spread about the center of the data (see Spread), the more uncertain the generalization. |
| Sampling variability | PTs are provided with samples of increasing size and are asked to generalize and to predict shape of a larger sample or of the population. | PTs are provided with samples of increasing size and are asked to generalize and to predict shape of a larger sample or of the population. | When the sample increases, PTs notice the s9-table features of distribution, such as the mean and mode, i.e., the shape of the sample distribution stabilizes. They will be confident that population distribution will have approximately the same characteristics as the sample distribution. | PTs identify features of the observed sample that are stable and features that are variable and use the stable features of sample distribution as arguments to generalize beyond the data. |
| Certainty | PTs are provided with samples of increasing size and with different distributions. | PTs are provided with samples of increasing size and with different distributions. | If the question to generalize is asked only once, PTs may not consider the certainty of their conclusions. Using samples with different sizes and distributions draws their attention to variation in the sample distributions. They will notice that their conclusion depends on the sample distribution and will conclude that their inferences are inherently uncertain. | PTs include probabilistic statements in their generalizations and compare the certainty of their conclusions for different sample size s. |

^aPTs stands for pre-service teachers

Table 9.4 Results of the analysis of the answer sheets

| Conclusion | Frequency |
|-------------------------------------|-----------|
| <i>Data as evidence</i> | |
| Yes | 36 |
| Other sources | 1 |
| No | 8 |
| Total | 45 |
| <i>Descriptive statistics</i> | |
| Global shape of distribution | 6 |
| Mean, median or sum | 8 |
| Spread | 14 |
| Local aspects of distribution | 2 |
| None | 18 |
| Total | 48 |
| <i>Type of conclusion</i> | |
| Descriptive | 1 |
| Unclear: Descriptive or inferential | 22 |
| Inferential | 9 |
| Probabilistic inferential | 1 |
| Refusal to generalize | 1 |
| None | 11 |
| Total | 45 |
| <i>Uncertainty</i> | |
| Probabilistic language | 3 |
| Sample size | 3 |
| None | 39 |
| Total | 45 |
| Prediction | Frequency |
| <i>Type of prediction</i> | |
| Shape smaller sample copied | 26 |
| Overemphasized conclusion | 9 |
| Shape smaller sample mimicked | 7 |
| Total | 42 |

the 45 conclusions, the participants used the data as evidence, although not always explicitly. For example, they wrote down only: “men are on average more positive.” We interpret the use of “average” as an indication that the average of a sample distribution was used as evidence. Only one group used another source as evidence in their conclusion. This group argued that women’s decreased logical and spatial thinking ability influenced their attitude toward mathematics.

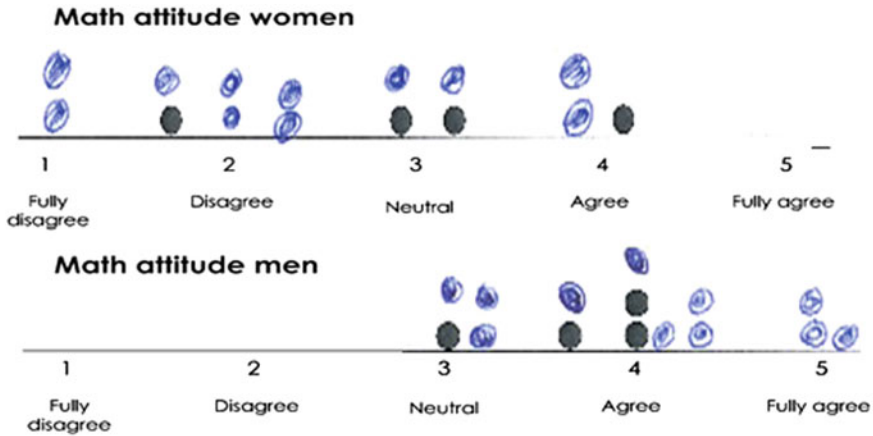


Fig. 9.2 Yasmine's and Esther's (Class B; both participants had a vocational education background) predictions of a larger sample that overemphasized the conclusion based on the small sample (pen writing constitutes the participants' predictions)

Concerning the descriptive statistics used as evidence, the participants often noticed the higher average of men and the high spread of women, but they did not connect the spread to their conclusion, as evidenced in this quote: "Many differences between men, few among women. Men love mathematics." Moreover, of the 34 answers that included conclusions only 18 were accompanied by a descriptive statistic as an argument, and only ten of these were supported by the mean or the global shape, which are suitable descriptive statistics to compare two distributions.

In the predictions, there were more indications that the participants used other sources of information. During the first round, six out of 11 groups overemphasized in their prediction the conclusion that men are more positive about mathematics than women, by moving the men's distribution to the right and the women's distribution to the left, as shown in Fig. 9.2.

Related to the learning objective Generalization, in 22 of the 34 answers that included a conclusion, it was unclear whether the conclusion pertained to the sample only or to the population. The following is a typical example: "Men are more positive about mathematics than women." Nine conclusions were coded as inferential because they included the words "in general." Only one group made a truly probabilistic generalization, by stating, "With more people, the conclusion we draw is more reliable, the attitude of men seems to be more positive."

Of the total of 42 predictions, 26 were plausible in the sense that they followed the global shape of the graph of the smaller samples. No group consistently smoothed the graphs, and only about half of the groups widened the range. Nine graphs mimicked the shape of the graphs of the smaller datasets, for example, by multiplying each frequency with the factor the sample increased, which is a very unlikely outcome (see Fig. 9.3). These results indicate that the participants did not understand that

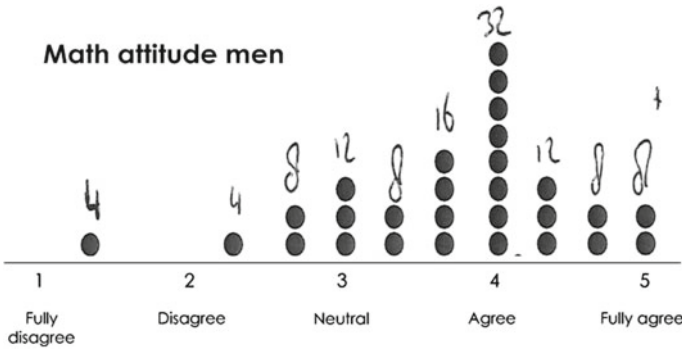


Fig. 9.3 Karel’s and Nick’s (Class C; other and vocational education background, respectively) predictions of a larger sample that mimics the smaller sample (pen writing constitutes the participants’ predictions)

a sample distribution resembles the population distribution more when the sample increases.

On the answer sheets, uncertainty was mentioned by only one group. This group concluded that no inference could be made based on four men and four women, and they were also the only group that made a probabilistic generalization. The other groups never mentioned uncertainty, apart from one weak indication of probabilistic language in the prediction of a graph (“We expect...”).

9.4.2 Class Discussions

Whereas the reasoning on the answer sheets was largely similar across the three classes, the class discussions differed in many respects. Therefore, we summarize these class discussions by class.

9.4.2.1 Discussion in Class A

In Class A, the first round started with several participants stating that male pre-service teachers are more positive about mathematics than female teachers. Cindy¹ objected, thus opening up the field to discuss ISI.

- Cindy: Here they are more positive, but just for these four persons. [...]
- Teacher educator: So you say the sample is too small.
- Cindy: It is just like you take four persons from a class and ask them what they think of the class. If you accidentally pick four positive

¹All the participants’ names are pseudonyms.

people from the class, you get a very good picture [inaudible], while the four other people could not like it at all.

Teacher educator: Ok, so we agree with this: the sample is just too small to say anything sensible about.

Various: Yes.

Cindy argued that another sample could result in an entirely different outcome because of the small sample size. She thus used ideas from sampling variability to explain why she thought that generalization was not possible. At first, all participants agreed. However, Merel objected by referring to the lower spread in the men's distribution: "Well, I think, the men's graph because it is less spread out, does say something, I think." She was very tentative about her conclusion that this lower spread about the center of the data meant more certainty about the population distribution compared to the higher spread in the women's distribution, but the teacher educator confirmed the validity of her reasoning. Next, the averages for men and women were compared. After the teacher educator pointed at the high spread about the center of the data in the women's graph, Merel again stated that the women's average was less informative because of the higher spread.

Turning to the predictions of the shapes of graphs for the 15 men and women, Yanka explained her prediction and indicated she had made the women's distribution more negative than the distribution of the smaller sample (see Fig. 9.2 for a similar prediction). Stressing the difference between men and women suggests she used other sources of evidence than the sample data for her prediction. A few participants suggested adjusting the predictions to make them resemble more the global shape of the women's graph for the sample of four.

During the class discussions in the second and third rounds, most of the time was spent on discussing how one could estimate the means without calculations, leaving little time to discuss ISI. During the second round, there was some attention to ISI when the teacher educator asked whether the participants expected the range to widen in the prediction for the larger sample. The participants thought it would but without showing indications of having thought about why the range would widen. It is thus doubtful whether the participants understood that the sample distribution resembles the population distribution more closely as the sample increases.

The third round, with samples of 28 men and 116 women, was concluded by summarizing the arguments used. The teacher educator asked about the role of spread regarding the conclusions. Initially, the participants struggled to answer this question, but finally, one said less spread meant more certainty about the conclusion. However, Merel, who introduced this claim during the first round, and Nicky appeared to doubt whether there were any stable signals in the sample distribution. The latter indicated that another sample (sampled the next week) may result in entirely different distribution:

Merel: We just don't know.

Nicky: [inaudible] next week it is on 1 again. [i.e., they are very negative about mathematics.]

In summary, at the start of the class discussion of the first round, Cindy's remark that a sample of four is too small for generalizations provided the opportunity to discuss ISI. This discussion yielded the insight that a sample of four is too small for generalizations. In addition, the group seemed to generally agree that the data could be used as evidence. Apart from these insights, little attention was paid to ISI. Moreover, understanding of how a sample can be used to generalize seemed to be absent.

9.4.2.2 Discussion in Class B

In Class B, the class discussions during the first and second rounds were short; moreover, only half of the participants participated. It was also unclear whether the conclusions were meant to be descriptive or inferential.

During the first round, the teacher educator asked for the participants' opinion about the reliability of the generalization. Manon responded that four teachers "never reveal the opinion of all pre-service teachers. One needs many more subjects." None of the participants supported or objected to this statement. Next, Manon described the graph of the larger sample of women, which she made more negative, using her own ideas about men's and women's attitudes about mathematics: "There are women who find it very difficult [...], and therefore, we made the women lower in the end." Again, the other participants did not participate in the discussion. During the second round, the group discussed the conclusion and the prediction only briefly.

During the third round, the discussion about the comparison of samples of unequal size yielded insights into the participants' conceptions of sampling variability. To compare the samples of 28 men and 116 women, some participants suggested multiplying the 28 men by 4 to make the sample sizes approximately equal (see Fig. 9.3). This deterministic approach, which neglects sampling variability, was challenged by other participants. For example, Rebecca stated, "Now you are estimating. You don't exactly know how men think and what they will fill in [...]. But in the end you can't know anything about it."

Later on, Yente explained why in her view this strategy is permissible:

Yente: Suppose at once there are all men who score 5 points. What we look at is, if it would go like this. How we think that it went exactly, then it is no problem. But we never know how other people think about it.

As Rebecca, Yente did not understand that a sample can be representative of the population. However, to solve the problem of comparing samples of unequal sizes, Yente simply *assumed* that if other men were sampled, they would have exactly the same sample distribution. She seemed to be primarily concerned with how to compare samples of unequal sizes, not with the issue of generalization:

Yente: If you just ask the Netherlands, then you can simply compare a number of men and women, but because among pre-service teachers there are not so many men, then you need to kind of estimate, I think.

Doubts about whether a sample can be representative of a population were also visible in the remarks of Marleen: “Yes, we can estimate. There are 116 women. [...] If there are suddenly ten more who all totally agree, then it is quite different from how it is now. It will always be estimation.” When challenged by the teacher educator, Marleen changed her mind:

Teacher educator: If there are ten more who all totally agree, you said. [Draws imaginary bar of ten women who all totally agree.] Is that likely?

Marleen: No.

Later on, she returned to her original idea:

Teacher educator: And Marleen says that suppose there are 15 women very positive, or ten, we immediately say that is very unlikely.

Marleen: It could still be.

Whereas Yente switched from complete uncertainty to complete certainty, by assuming that other men had the same attitude as the men sampled, Marleen remained doubtful about what result one could expect in a different sample. Only Rebecca seemed to believe in the possibility of making generalizations: “That is the way they always do it, right? If they want to know something, they don’t ask [inaudible; probably: everyone].”

Overall, except for Rebecca, the Class B participants were reluctant to accept the claim that a sample can be representative of a population. The class discussions showed that this reluctance was probably caused by their lack of understanding of sampling variability.

9.4.2.3 Discussion in Class C

In Class C, the first round was skipped because during the third round in Class A and B motivation seemed to decline. The second round started with the conclusion that men are more positive about mathematics than women. Whether this conclusion was meant to be descriptive or inferential initially was unclear. In an extensive discussion about the use of measures of centrality as arguments, the group concluded that neither the midpoint of the range nor the mode for non-normal distributions is useful in comparing distributions. Next, seven of the nine participants participated in a lively discussion about whether the conclusion would hold for the population. Initially, various participants denied this.

Teacher educator: If we look at these men... I am curious to know, can we say something about those 500 based on 15 men [500 was the assumed size of the men’s population]?

Khlood: But this is not a good sample, is it?

Teacher educator: Why is this sample not good?

Khlood: Because there are way too few.

Later on, Khloud related the size of the sample to sampling variability: “Yes, at random, I understand, but maybe you by chance picked the best 15.” Two other participants agreed with this statement, and one concluded that “in the end, one can never say something about it,” which multiple participants agreed with. Then the teacher educator asked about the relationship between spread and certainty.

Teacher educator: About whose attitude do you have more certainty: about men or women?

(Almost) all: Men.

[...]

Karel: The chance is higher that they are all over there [points at the positive part of the graph] than they are not over there.

Karel claimed that it was more likely that another sample of men would also be predominantly positive about mathematics. Various participants agreed with Karel. The teacher educator confronted the participants with their conflicting opinions: Earlier they had said that nothing could be known, but then they indicated they had more certainty about men. Khloud’s response illustrates how the participants appeared to combine these opinions: “We know the chance, but we are not sure.” Laura concluded, “Because you did such a small sample, [inaudible] you never know for sure [inaudible] I still don’t think it is a good sample.” The participants wanted not only to have a larger sample but also seemed to want a sample that would give them complete certainty about their generalizations. Making generalizations with a certain degree of uncertainty appeared to be problematic for them.

During the third round, ways to compare samples of unequal sizes were discussed. Although, as in Class B, some participants suggested multiplying the 28 men by four (see Fig. 9.3), or, alternatively, dividing the 116 women by 4, three participants argued that one could use the mean to compare samples of unequal sizes because the shape of a sample is expected to remain approximately the same as the sample increases. In this discussion, Khloud showed a remarkably good understanding of the effect of sample size on sample-to-sample variability: “In any case, if you would take 15 people, then the chance is smaller it remains the same [...] than if you have 28 people.” Overall, a small majority of the participants indicated that they expected approximately the same results for a different sample, while none of the participants expressed the opposite.

In sum, the majority of the Class C participants seemed to understand that it is possible to make generalizations because the shape of a sample is expected to remain approximately the same as the sample increases. Comparing the spread and the certainty about the two distributions led the participants to express correct ideas about sampling variability. However, they displayed an inclination to demand complete certainty about the generalizations.

Table 9.5 Quality of reasoning about ISI per class

| ISI component | Aspect | Class | | |
|---------------------------------------|----------------------|-------|-----|----|
| | | A | B | C |
| <i>Data as evidence</i> | Data as evidence | ++ | + | ++ |
| | Center | + | – | + |
| | Spread | + | – | 0 |
| | Distribution | 0 | 0 | 0 |
| <i>Generalization beyond the data</i> | Generalization | 0 | – – | + |
| | Prediction | 0 | – | – |
| <i>Probabilistic language</i> | Sample size | – | – – | – |
| | Heterogeneity | – | – – | + |
| | Sampling variability | – | – – | 0 |
| | Certainty | 0 | – – | + |

Note Quality of the reasoning about ISI, ranging from – – (*reasoning not at all in line with learning objective*) to ++ (*reasoning entirely in line with learning objective*). See Table 9.2 for a detailed explanation of the indicators

9.4.3 Quality of Reasoning About ISI

Table 9.5 shows the quality of reasoning about ISI for each class. Indicators were assigned to the learning objectives that showed to what extent the pre-service teachers attained the learning objective (see Table 9.2). For most aspects, Class C’s reasoning about ISI was the most sophisticated, although the first round was skipped in this class. Class B’s reasoning was poor overall.

In all three classes, elaborate discussions about estimating the mean took place. To most participants, it seemed clear that a comparison of the means or the global shape of the sample distributions was a valid way to compare the distributions. However, although many participants noticed the high spread about the center of the data in the women’s distribution, few understood that this made a generalization more uncertain. In Class A, two participants discussed the effect of the spread on the choice of the measure of centrality.

On the answer sheets, there was little evidence that the participants intentionally generalized to the population. Questioning by the teacher educator during the class discussions made generalization a topic of discussion. To accept the feasibility of making generalizations, an understanding of sampling variability proved vital. In Class B, all but one of the participants thought nothing could be known about the population as a whole, since another sample could differ significantly from the current sample. In Classes A and C, most participants seemed to understand that making uncertain generalizations is possible. Class A’s reasoning was superficial. Most of the participants acknowledged the uncertainty of generalizations and that a sample of

four is too small for any generalization, but what could be stated about a population based on a sample was not discussed. Class C's understanding of generalizations was more sophisticated. They agreed that less spread means more certainty. However, they demanded complete certainty about the generalizations, dismissing samples that could not provide this complete certainty.

The participants' predictions of the distributions of larger samples and the population revealed many did not understand that sample distributions would resemble the population distribution when the sample increases. Classes B and C made many predictions that exactly copied the shape of the distribution of the smaller sample. In Class C, discussing the predictions was only a small part of the discussions, which might partly explain the low quality of reasoning about predictions in this class.

Understanding of the learning objectives of the Probabilistic language component differed among the three classes. Even in the best-performing class, Class C, the ISI reasoning was not as sophisticated as expected. In all classes, the answer sheets almost completely lacked attention to uncertainty. Class B's reasoning about uncertainty was the least developed. Although there was broad consensus that for the sample of four generalization was impossible, other uncertainty aspects were not discussed. The participants did not use probabilistic language, except for the opposite, certainty language, for example, in statements, such as "It can never be the case..." In addition, the Class A participants agreed that for the sample of four, generalization is impossible. Moreover, the majority of this class seemed to understand the possibility of making uncertain generalizations. Only Class C had an intense class discussion about the extent to which a sample may provide information about other people not interviewed. Even there, however, the majority was not convinced that a larger sample would look like the smaller sample. Only a minority could express the idea that the shape of a sample is likely to remain approximately the same when the sample increases, provided the sample is sufficiently large. The participants in Class C regularly used uncertainty language during the class discussions.

9.5 Conclusions and Discussion

This study explored the growing samples heuristic in the context of teacher education. We investigated how three classes of first-year pre-service primary education teachers reasoned about ISI when engaged in a growing samples activity. The results show that in two classes most seemed to agree that making (uncertain) generalizations based on a sample is possible. However, overall, the majority was unable to link the possibility of making generalizations to an understanding of how a well-selected sample can be representative of the population.

Concerning the way descriptive statistics were used as arguments in ISI, the class discussions revealed that most pre-service teachers implicitly used suitable descriptive statistics to compare two distributions. On the answer sheets, however, only a third of the conclusions were supported by suitable descriptive statistics. In particular in Rounds 1 and 2 it could easily be seen, without calculation, that men were on aver-

age more positive about mathematics than women. The difference may have been too obvious to induce the pre-service teachers to write down descriptive statistics.

On the answer sheets, most conclusions seemed to describe the sample data only, rather than generalize beyond the sample data. Inferential statements used at best the colloquial term ‘in general,’ which could have been copied from the question without the intention to generalize. The first explanation for this finding may be that the need to generalize was not compelling enough. Another explanation could be that the participants, in their role as future teachers, had a class of primary school students in mind as their population of interest. When the class is the population of interest, description suffices, and there is less inclination to generalize beyond this class. The questioning by the teacher educator during the class discussions was necessary to draw the pre-service teachers’ attention to the inferential nature of the questions.

Attention to uncertainty and sample size was virtually absent on the participants’ answer sheets. This underlines our conclusion that most of the pre-service teachers described only the sample data. Description does not require uncertainty and sample size to be taken into account. During the class discussion in Class B, the majority of the pre-service teachers concluded that generalization is impossible because they accepted the claim that nothing can be known about people who are not in the sample. This resembles the instance found by Ben-Zvi et al. (2012) of students uttering complete uncertainty. In Classes A and C, in contrast, the majority of the participants seemed to acknowledge that making uncertain generalizations based on a sample is possible. This finding is similar to the findings in De Vetten et al. (2018) and the reasoning displayed by high-ability middle-grade students studied by Ben-Zvi et al. (2012).

We found little evidence that the heuristics helped the pre-service teachers to understand the concept of sampling variability, contrary to the ideas formulated by Joan Garfield and Ben-Zvi (2008). Only one participant attempted to explain the stability of sample distributions when the sample increases by referring to probability theory. Her explanation did not convince the other pre-service teachers. The predictions of the distribution of larger samples and the population provided extra evidence for this finding, since often these predictions too strictly followed the global shape of the sample at hand. Not understanding sampling variability is problematic because it seemed to make the participants reluctant to accept the possibility of making generalizations. One reason why the activity did not foster an understanding of sampling variability could be that for a given sample size, all groups received the same data set. Understanding why the sample distributions become stable when the sample increases might require a repeated samples approach, where each group receives a different data set and compares their conclusions with other groups.

In Class C, only the second and third rounds were used, but their generalization and sampling variability were the most sophisticated. The questioning by the teacher educator appeared a more effective way to foster reasoning about these topics than repeatedly asking the participants to generalize and make predictions. The results also raise questions related to the optimal number of rounds, the sample size of the first round, and the effect.

These results show some benefits of the growing samples heuristic in general and our operationalization in particular. First, the heuristic helped to initiate discussions about the role of sample size in certainty and sampling variability, which are key concepts in ISI. In addition, using sample distributions with different variabilities seemed to have helped the participants to gain insight into the certainty of generalizations. Second, the activity was useful for discussing many distributional aspects, such as measures of centrality and the effect of spread on measures of centrality, as was the case in Bakker's (2004) study. However, discussing the calculation and estimation of descriptive statistics took considerable class time, which could have been spent more productively on how one can use descriptive statistics as arguments in ISI, for example, what the spread of the sample distributions implies for the conclusions. Third, the use of samples of unequal sizes during the third round initiated discussions about using the measure of centrality in comparing different sample sizes and about sampling variability.

The participants' educational background could have played a role in the different quality of reasoning about ISI between the classes. This background is clearly different for Class B, compared to the other two classes. In Class B, nine of the 11 participants have a background in secondary vocational education but only seven out of 15 in Class A and four out of eight in Class C. Since statistics is not part of most secondary vocational education curricula but is part of most senior general secondary education curricula, the pre-service teachers with a background in senior general secondary education may have had the vocabulary and statistical tools to further the reasoning about ISI during the class discussions. We found some evidence for this explanation. One pre-service teacher in Class C explicitly stated, "I had something about this in secondary school." Moreover, one pre-service teacher with a background in senior general secondary education introduced the term "probability theory", after which probability and chance became terms used to reason about sampling theory. For a fruitful ISI discussion, a fair number of pre-service teachers with appropriate background knowledge in statistics and probability theory may need to be present.

Some issues warrant a cautious interpretation of the results. First, this was a small-scale and explorative study, and the context was the Dutch educational system where students enter teacher college immediately after secondary education. The results are, accordingly, not readily generalizable to other contexts. However, similar processes may occur in countries where students enter teacher college with similar backgrounds and where the statistics curriculum in primary and secondary education is comparable to the Dutch system. Second, the design of the activity likely influenced the reasoning. For example, the sample distributions may have influenced the results. In particular, the data did not result in conflicting conclusions. During each round, it was quite obvious that men were more positive about mathematics than women. Third, sound recordings of the pre-service teachers when working in groups were not available. Issues spoken about but not written down could have provided a more complete picture of the participants' reasoning, in particular about whether they had spoken about generalization and uncertainty but had not written down these issues. Nonetheless, the extensive class discussions of generalization and uncertainty

probably provide a reliable general impression of the pre-service teachers' reasoning about ISI.

In conclusion, this study informs how the effectiveness of the heuristic can be further strengthened. First, the pre-service teachers seemed to use correct descriptive statistics as arguments in ISI. This finding indicates less focus might be given to descriptive statistics and by using simple descriptive statistics, more on ISI itself. Second, some pre-service teachers were reluctant to accept the possibility of making generalizations beyond the data. Comprehension of this fundamental idea may be fostered if each group uses a different data set. When the sample sizes increase, the different data sets typically will begin to resemble each other, leading to confidence on the learner's behalf that from a certain sample size onward, a sample provides reliable information about the population. Finally, because the pre-service teachers tended to describe the data only, the need to make generalizations beyond the data was not sufficiently compelling. Therefore, we recommend designing activities and contexts in which description is clearly insufficient and where generalization beyond the data is natural and inevitable. These changes to the growing samples heuristic may help to provide pre-service teachers the knowledge to demonstrate to primary school students the feasibility of making generalizations beyond the data.

References

- Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: CD-β Press, Center for Science and Mathematics Education.
- Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, 13(2), 5–26.
- 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.
- Batanero, C., & Díaz, C. (2010). Training teachers to teach statistics: What can we learn from research? *Statistique et enseignement*, 1(1), 5–20.
- Ben-Zvi, D. (2006). *Scaffolding students' informal inference and argumentation*. Paper presented at the Seventh International Conference on Teaching Statistics, Salvador, Brazil.
- Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM—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.
- Ben-Zvi, D., Gil, E., & Apel, N. (2007). What is hidden beyond the data? Young students reason and argue about some wider universe. In D. Pratt & J. Ainley (Eds.), *Proceedings of the Fifth International Forum for Research on Statistical Reasoning, Thinking and Literacy (SRTL-5)*. Warwick, UK: University of Warwick.
- Burgess, T. (2009). Teacher knowledge and statistics: What types of knowledge are used in the primary classroom? *Montana Mathematics Enthusiast*, 6(1&2), 3–24.
- Canada, D., & Ciancetta, M. (2007). *Elementary preservice teachers' informal conceptions of distribution*. Paper presented at the 29th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Stateline, NV.

- Cobb, P., & Tzou, C. (2009). Supporting students' learning about data creation. In W.-M. Roth (Ed.), *Mathematical representation at the interface of body and culture* (pp. 135–171). Charlotte, NC: IAP.
- De Vetten, A., Schoonenboom, J., Keijzer, R., & Van Oers, B. (2018). Pre-service primary school teachers' knowledge of informal statistical inference. *Journal of Mathematics Teacher Education*. <https://doi.org/10.1007/s10857-018-9403-9>.
- 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., 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.
- Groth, R. E., & Bergner, J. A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning*, 8(1), 37–63.
- Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. 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. 235–246). Dordrecht, The Netherlands: Springer.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., et al. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction*, 26(4), 430–511.
- Jacobbe, T., & Carvalho, C. (2011). Teachers' understanding of averages. 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. 199–209). Dordrecht, The Netherlands: Springer.
- 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.
- Leavy, A. M. (2006). Using data comparison to support a focus on distribution: Examining preservice teacher's understandings of distribution when engaged in statistical inquiry. *Statistics Education Research Journal*, 5(2), 89–114.
- Leavy, A. M. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning. *Statistics Education Research Journal*, 9(1), 46–67.
- Liu, Y., & Grusky, D. B. (2013). The payoff to skill in the third industrial revolution. *American Journal of Sociology*, 118(5), 1330–1374.
- 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.
- Makar, K., & Rubin, A. (2014). *Informal statistical inference revisited*. Paper presented at the Ninth International Conference on Teaching Statistics (ICOTS 9), Flagstaff, AZ.
- Meletioui-Mavrotheris, M., Kleanthous, I., & Paparistodemou, E. (2014). *Developing pre-service teachers' technological pedagogical content knowledge (TPACK) of sampling*. Paper presented at the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, AZ.
- Meletioui-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.
- Mooney, E., Duni, D., VanMeenen, E., & Langrall, C. (2014). Preservice teachers' awareness of variability. In K. Makar, B. De Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9)*. Voorburg, The Netherlands: International Statistical Institute.

- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. London, UK: Temple Smith.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- 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.
- 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 10

Necessary Knowledge for Teaching Statistics: Example of the Concept of Variability



Sylvain Vermette and Annie Savard

Abstract This chapter explores teachers' statistical knowledge in relation to the concept of variability. Twelve high school mathematics teachers were asked to respond to scenarios describing students' strategies, solutions, and misconceptions when presented with a task based on the concept of variability. The teachers' responses primarily helped us analyze their comprehension and practices associated with the concept of variability and gain insight into how to teach this concept. Secondly, the study shows that students and high school teachers share the same conceptions on this subject.

Keywords Professional knowledge · Statistics · Teacher's knowledge
Teaching practices · Variability

10.1 Context

The importance of statistics in our lives is such that data management has become a major key in the education of responsible citizens (Baillargeon 2005; Konold and Higgins 2003). The abundance of statistical data available on the internet, the studies reported on television news, or the studies and survey results published in newspapers and magazines all show that nowadays, citizens must have analytical skills to develop critical judgment and a personal assessment of the data they are confronted with daily. This role of statistics in our current society makes it necessary to consider teaching this discipline to train our students to be citizens of tomorrow. If the goal is to encourage statistical thinking in students as future citizens, then not only do we need to teach basic statistical data interpretation skills, but it is also essential to teach the

S. Vermette (✉)

Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada

e-mail: sylvain.vermette@uqtr.ca

A. Savard

McGill University, Montréal, Québec, Canada

e-mail: annie.savard@mcgill.ca

© Springer Nature Switzerland AG 2019

G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_10

225

concept of variability. Variability is a key concept for the development of statistical thinking; statistics may be defined as the science of the variability in natural and social events in the world around us (Wozniak 2005).

We live in a world characterized by variability. Take the example a business that manages a city's public transit system. It may announce that its trains will arrive at the different stations every ten minutes. However, any regular transit user knows that arrival times vary, and that the time tables are not always strictly respected. The time intervals are unequal, and this lack of uniformity is a characteristic of the presence of variability. Moreover, the variable number of travellers must also be considered. This variable reflects the somewhat predictable variation of factors such as schedules or seasons and the random and unavoidable daily variability for a given hour. In short, as shown in this one of many possible examples, variability is reflected by the absence of determinism. The complexity of the phenomena, affected particularly by the number of variables involved, is the source of this variability and of the observed variations. In the public transit example, studying the phenomenon in all its variability makes it possible to ensure a generally satisfying and consistent service by planning the required train capacity and anticipating a variable but reasonable delay between train runs.

Acknowledging an event's variability means recognizing that the results are subject to variation, understanding that they are unpredictable in the short run, considering the sampling fluctuations, and letting go of certainty to enter the world of uncertainty (Vergne 2004). In statistics, the concept of variability might be thought of as having two dimensions: sample fluctuations in different samples taken from a same population and dispersion of data in a distribution that can be assessed using measures of spread. This last dimension is the focus of the present article. In Quebec, curriculum documents (Ministry of Education and Higher Education 2004, 2007) introduce and stress early ideas associated with variability related to contexts of sampling, particularly in a probabilistic context. The curriculum also introduces ideas associated with variability related to statistical data dispersion, for example measures of dispersion (range, mean absolute deviation, standard deviation), graphing and exploratory data analysis. So, this statistical key concept is implicitly part of the academic curriculum in school but without being clearly defined. This is problematic because in everyday language, this concept may be understood as large variety being associated with large variability. This common idea of variability as a measure of variety differs from its statistical concept associated with concepts of variation and variance. As stated by Reading and Shaughnessy (2004), variability is the cognitive processes involved in describing both the propensity for change and measure of variation. The concept of variability is inevitably associated with the concept of variation. As opposed to Borim da Sina and Coutinho (2008) and Reading and Shaughnessy (2004), we do not make a distinction between variation and variability in this book chapter.

Based on the foregoing, it is essential to teach the concept of variability to develop students' statistical thinking. It is therefore necessary to verify the teachers' knowledge on this subject because we assume that students' knowledge development is closely linked to the practices and knowledge of their teachers who support them and organize teaching by creating environments conducive to learning. We may think that

understanding conceptions related to a specific concept helps teachers not only to better plan their teaching but also to better organize and manage students' activities in a classroom so they can learn the elements of a targeted mathematical concept. We define conceptions as explicative models based on daily life experience for explaining and interpreting phenomenon, as is, to make sense of the world (Savard 2014). We consider them as personal knowledge; however, for the purpose of this paper, we call them misconceptions.

Despite the presence of statistics and probability in the school curriculum in Quebec, training in statistics and in teaching this discipline is barely present in the universities' teacher training curriculum despite the ever-growing presence of statistics in academic programs (Gattuso and Vermette 2013). For example, our study shows that, unlike for geometry or algebra, no class is exclusively dedicated to the didactics of statistics in the Quebec universities teacher training programs. One may think that mathematics teacher training in statistics and in teaching statistical concepts is not very developed, which raises important questions about the nature of the statistical experience encountered by teachers during their professional training. Several studies show the growing interest of high school mathematics teachers' in understanding the statistical concepts they teach (e.g., Bargagliotti et al. 2014; Dabos 2011; Garfield et al. 2007; Green and Blankenship 2014; Hill et al. 2005; Silva and Coutinho 2006). At the same time, we notice a growing awareness that teachers use specific forms of knowledge within their practice that are different from the standard forms they learned in their university mathematics courses (Moreira and David 2005, 2008; Proulx and Bednarz 2010, 2011; Margolinas 2014). Recent developments related to teachers' mathematical knowledge show that some knowledge stems from teaching practice and is therefore related to events from the learning/teaching context (Bednarz and Proulx 2009; Davis and Simmt 2006; Margolinas 2014). This interaction between statistical training and practice in class is the central issue of the research project described in this chapter.

Learning more about teachers' knowledge related to the concept of variability might provide a better understanding of their ability to teach this concept. The research question in this chapter is: What is the high school teachers' mathematical knowledge about the concept of variability? In the following, theoretical anchors around the concept of variability in statistics that guide this study are introduced along with a clarification of what teachers' professional knowledge means. After considering the methodological aspects of the study, the analysis offers examples of teachers' strategies associated with tasks involving the concept of variability. The chapter concludes with a discussion of the results with the perspective of training future teachers.

10.2 Theoretical Anchors

10.2.1 *The Concept of Variability*

One focus of statistics is inferring characteristics of a population through analysis of data collected from a sample of the population. Here, the concept of variability appears in the differences between different samples taken from a same population. Understanding this variability is required to make statistical inferences because such inferences essentially come with probabilistic uncertainty caused by sample fluctuation. The concept of variability is then also linked to the development of probabilistic reasoning as it allows, in a mathematical context, to forsake a determinist reasoning, and reason instead using the uncertainty caused by these sample fluctuations. Some researchers such as Canada (2006) studied teachers' interactions with this concept. Others studied teachers' understanding of variability: for example, Dabos (2011) studied teachers' conceptions; Peters (2011, 2014) studied the development of this concept by teachers; Sanchez et al. (2011) studied teachers understanding of variation.

Another important focus of statistics is interpreting descriptive statistics. As already stated, we consider that the concept of statistical variability refers to the dispersion of statistical data. A data distribution shows variation, and, although information on an important dimension of a distribution is provided by measures of central tendency, used alone such measures may suggest an incomplete representation of the distribution. We then need to pay attention to the variability of the statistical variable's values, which can mainly be evaluated with dispersion measures that show the variation of data in a distribution.

A measure of dispersion allows a data set to be described by a specific variable that provides an indication of the variability of the values within the data set (Dodge 1993, p. 225, translated from French). A widely-used dispersion measure for describing a distribution's variability is the range. It is not only used because it is easy to calculate simply by obtaining the difference between the largest and the smallest value of a distribution but also because the results are easy to interpret (the size of the smallest interval that contains all the data). Used alone, the range is a limited way to measure variability as it does not consider the influence of the frequency of each statistical variable's values on the variability. Another way to measure variability is to use the mean and standard deviations, two measures that include all the distribution's data, and describe the data dispersion around the distribution's center, in other words, the distribution's mean. Interpreting these measures is a bigger challenge. Recent studies show that understanding these statistical measures is often limited to the calculation of algorithms and thus highlight students' difficulties in measuring variability in terms of the proximity of the data to the distribution's central point (Cooper and Shore 2010; Dabos 2011; delMas and Liu 2005; Meletiou-Mavrotheris and Lee 2005). This seems to be particularly true with graphical representations; it seems challenging to make connections between informal notions of variability based on graphical displays and more formal measures of variability (Cooper and Shore 2010).

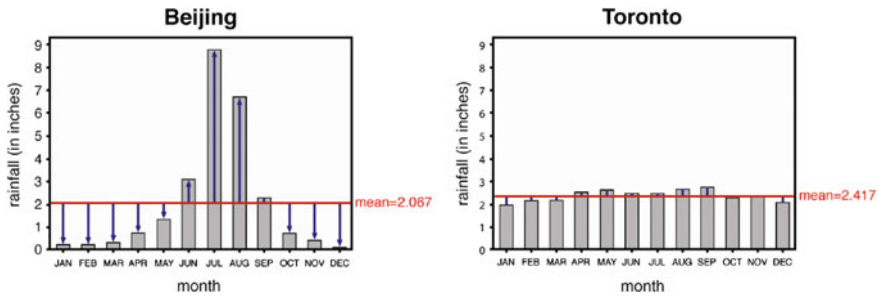


Fig. 10.1 Illustration of variability with vertical lines showing deviations from the mean (Cooper and Shore 2010, p. 5)

According to Garfield and Ben-Zvi (2005), being able to recognize and understand how the concept of variability appears in different graphs, especially in histograms, is an important aspect for developing this concept when we consider that graphs’ appearance is an obstacle that may induce alternative conceptions. This aspect seems overlooked in high school where too often the focus is rather on the rules for creating the graphs. Students may take in the general look of a graph, the maximum, minimum or outliers and still not deeply understand the relationships between the distribution’s center and the data spread around the center leading to the concepts of mean and standard deviation as ways to study variability (Cooper and Shore 2008).

Several misconceptions interfere with this key reasoning on the variability of a statistical variable’s values, which is interpreting the data spread in terms of deviation from the distribution’s center, when dealing with variability in histograms and bar graphs. Variability may first be interpreted as a variation of the height of bars and conclude that the wider variance in the heights of bars, the greater the variability (Cooper and Shore 2008; Meletioui-Mavrotheris and Lee 2005). Reasoning based on the difference between the bar heights leads to thinking about the problem vertically (if the bars are vertical). This understanding of variability might lead to an incorrect answer if the value corresponding to the mean is on the x-axis bearing the different values of the variable. The deviations from the mean might be illustrated with horizontal lines. These deviations could certainly have also been represented by vertical lines if the different values of the variable, and therefore the mean, had been on the y-axis as shown in Fig. 10.1.

These two graphs illustrate the mean monthly amounts of rainfall in Beijing and in Toronto. The values of the variable, the mean amount of rainfall per month, are on the y-axis and do not include the frequency. It is therefore possible to compare the variability in the mean amounts of rainfall by observing the difference between the heights of bars and the line representing the overall mean amount for the year and to conclude that there is a smaller variability in the mean monthly rainfall in Toronto than in Beijing. In this case, the deviations from the mean may be illustrated by vertical lines corresponding to the differences per month of the mean amount of rainfall from the yearly mean rainfall. In short, a distribution’s variability cannot be

judged based on the variation in the height of bars but rather on how the vertical segments cluster around the yearly mean. Some graph types, as in Fig. 10.1, which do not show the distribution of frequencies, may be problematic and difficult to interpret.

Other elements associated with the visual aspect of the shape of the distribution might influence the interpretation of variability in a graphical representation. For example, one element is to use the Gaussian or normal curve as a reference to compare a graphical representation. When the graphical representation looks like a Gaussian curve, the variability of the distribution is perceived as low. This way, variability is seen as a deviation from normalcy (Dabos 2011). Another example is associating a symmetrical distribution with low variability (delMas and Liu 2005; Meletiou-Mavrotheris and Lee 2005). In this case, one may believe that this association is influenced by the distribution's shape that shows a perfect counterbalance of the deviations of the statistical variable's values from the center of the distribution. Students might be ignoring the fact that the sum of all of the deviations from the mean is the null and that to find a measure of dispersion around the mean, it is necessary to use some technique such as the absolute value of each deviation, so the negative deviations don't offset the positive deviations.

Other aspects related to the variability of the values of a statistical variable may certainly be documented. However, descriptive statistics and graphical interpretations are the preferred aspects for building teachers' tasks (see Sect. 10.3). Another aspect, related this time to the knowledge mobilized by the teachers in their practice, was also put forward: the professional knowledge.

10.2.2 Professional Knowledge: Knowledge Linked to Practice

Inspired by the works of Shulman (1986, 1988), two aspects of teacher knowledge present a specific interest: content knowledge and pedagogical content knowledge. Shulman defines content knowledge as how a specialist in a specific field understands a related subject matter. Pedagogical content knowledge is the ability to organize and manage students' activities in the classroom, so they may be introduced to the elements of a targeted mathematical knowledge (Bloch 2009). Pedagogical content knowledge is the ability to introduce and explain a topic for others to understand. This type of knowledge goes beyond content knowledge and focuses on a different dimension; understanding the content to teach it (Holm and Kajander 2012; Proulx 2008).

It is possible to separate these two types of knowledge; however, in practice, they are interrelated and very hard to distinguish (Even 1993; Even and Tirosh 1995). Therefore, this study does not seek to distinguish content knowledge from pedagogical content knowledge but rather focuses on the conceptualization of professional mathematics based on the works of Moreira and David (2005, 2008), of Proulx and

Bednarz (2010, 2011) and of Margolinas (2014) who consider academic and school mathematics as two separate knowledge fields. For example, in teaching/learning mathematical situations, many mathematical aspects arise and are considered by the teacher: reasoning (appropriate or not), giving meaning to the concepts; conceptions, difficulties and errors on the comprehension of concepts; various strategies and approaches to solve a problem; various representations, symbols/writings (standardized or not) to express solutions; and new questions and paths to explore. These mathematical occurrences not only refer to concepts in curricular documents, which dictate what must be taught, but also refer to mathematical elements that are part of teaching/learning mathematics that the teacher must use in class. The teacher's professional mathematical knowledge refers to a body of knowledge and mathematical practices built on teaching/learning mathematics (Bednarz and Proulx 2010).

This mathematical orientation based on practice (Even 1993; Even and Tirosh 1995) is at the heart of the present research. Here, high school mathematics teachers' professional knowledge is studied from two perspectives based on tasks involving statistical content and related students' reasoning (see Sect. 10.3). The first perspective refers to the teachers' knowledge of the concept of variability. Are the teachers able to perform the task and identify the students' misconceptions? The second perspective is their ability to intervene with students to help them reason from their errors.

10.3 Methodology

This exploratory project, which was conducted in French, is part of a larger research program focused on issues associated with teaching statistics with the objective of developing and analyzing training to improve the statistical experience. To achieve the present research's goal, which is to learn more related to mathematics high school teachers' statistical professional knowledge about the concept of variability, we interviewed teachers using scenarios involving the concept of variability to collect data from teachers' answers and to better understand their ability to teach this concept. Twelve Quebec high school mathematics teachers participated in the research project, and all had studied one course in statistics in their initial training to be a mathematics teacher. All of them had a minimum of five years experience in teaching mathematics in secondary school. Since interviews were conducted at the end of the school year, these teachers had already taught a chapter on statistics to their students during the year.

It was a two-step experiment. First, an information letter was sent to teachers coming from a school district, inviting them to participate in the research project. This letter also briefly introduced the concept of variability and the purpose of the study. Introducing the concept of variability was necessary since it is not expressly defined in the Quebec school curriculum. The following statements were presented:

- the aim of the study is to explore how the concept of variability is considered while teaching,
- the concept of variability refers, among other things, to the dispersion of data in a distribution and to sample fluctuations and,
- the possibility of quantifying the variability in a distribution of data by using dispersion measures such as the range, interquartile range and standard deviation.

Second, when a teacher agreed to participate in the research project, he or she was invited to meet the principal investigator for an interview, which consisted of analyzing six scenarios. Twelve teachers were interested and were thus selected to participate. The interviews were conducted in a high school situated in the middle of the school district. In each of the scenarios, a statistical problem was presented (*Problem solving*), along with one or two students' solutions (*Students' responses to the problem and interventions*). These were built using statistical contents analyses related to the concept of variability (didactical, conceptual, and epistemological analysis; Brousseau 1998) and inspired by analyses performed in this field (Cooper and Shore 2008, 2010; Dabos 2011; delMas and Liu 2005; Meletiou-Mavrotheris and Lee 2005). For each of the scenarios, teachers had to make sense of the task, of students' thinking, and propose possible interventions to improve students' statistical reasoning and understanding. This process provided information on the teachers' professional knowledge of the concept of variability. In this chapter, we present two scenarios involving a graphical representation linked to variability. The first scenario focused on a distribution of data, while the second scenario focused on standard deviation.

10.3.1 First Case Example¹

Step 1: Problem solving

The charts below show the height, in centimeters, of Secondary 1 (grade 7) students from two different schools. Each school has 93 students. Which chart shows the greatest variability in the students' height? Explain your choice (Figs. 10.2 and 10.3).

Step 2: Students' responses to the problem and interventions

Although they reasoned differently, two students came to the same conclusion for this question; school B's graph shows a greater variability. The first student's reasoning is that school B's graph has an oscillating pattern. The second student finds school A's graph almost symmetrical and concludes that school B's graph shows a greater variability. What do you think of the students' answers? Which reasoning do you favor? How would you respond to each student?

¹Adapted from Canada (2004).

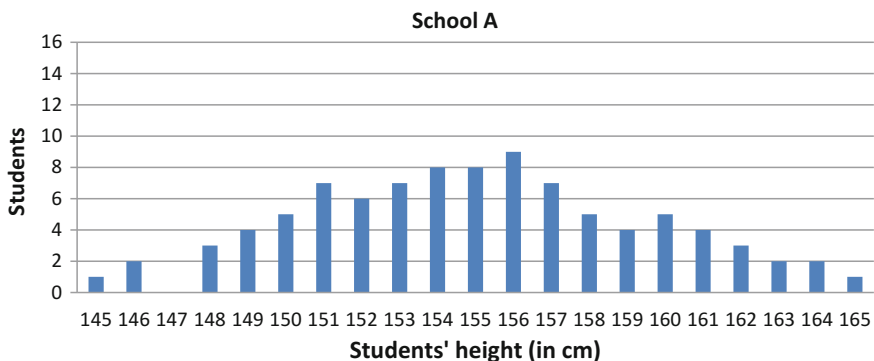


Fig. 10.2 Height, in centimeters, of Secondary 1 (grade 7) students from school A

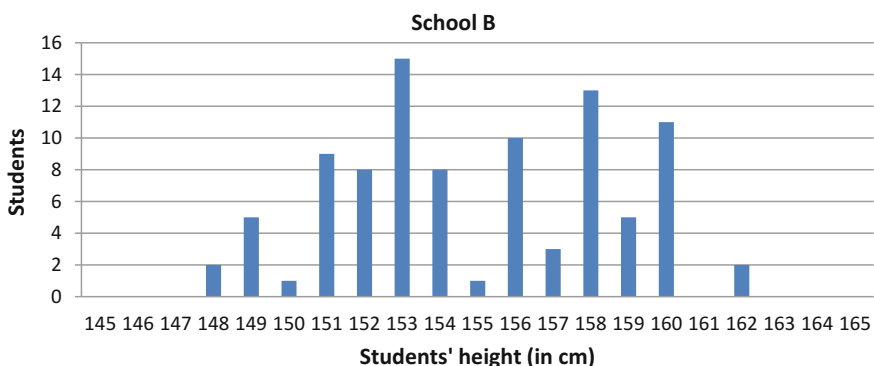


Fig. 10.3 Height, in centimeters, of Secondary 1 (grade 7) students from school B

In this scenario, the concept of variability is shown in the way the data are dispersed in the two distributions of the heights. The correct answer is: School A has the distribution with the greatest variability in the students' height. The task given to the teacher stems from the students' misconceptions. The students' reasoning highlighted in this problem are based on the works of Cooper and Shore (2008), delMas and Liu (2005) and Meletiou-Mavrotheris and Lee (2005), who reported that, when interpreting the variability of a data distribution, some students were influenced by aspects related to the shape of the distribution. The first student's answer is designed to be influenced by the variation in height of school B's bars. This student's reasoning refers to the frequency variability and not to the variability in the subjects' heights. The second student's response is designed to be influenced by the symmetry of school A's chart.

10.3.2 *Second Case Example*²

Step 1: Problem solving

Throughout the year, a teacher collected statistics on the quantity of water drunk by Secondary 4 (grade 10) students from her school. The school has three groups each with 27 Secondary 4 students. The statistics she collected are in Fig. 10.4 here below. When looking at groups A, B and C, which distribution has the greatest standard deviation? Which distribution has the smallest standard deviation? Explain your choice.

Step 2: Students' responses to the problem and interventions

For this question, two students came to different conclusions regarding group C. The first one said that group C's graphical representation shows the distribution with the largest standard deviation since it has the most bars which indicates a large variety of the quantity of water drunk monthly by students. He then concluded that this distribution has the largest standard deviation. The second student said that the graphical representation of group C shows the distribution with the smallest standard deviation. He based his reasoning on the fact that the bars from group C's graphical representation are at an even height and therefore this distribution has the smallest standard deviation. Who is right? How will you respond to these students?

In this case example, the concept of standard deviation can be seen through the dispersion of the quantities of water in the histograms representing the two distributions. The correct answer is: Group A has the distribution with the smallest standard deviation, and Group B has the largest standard deviation. Again, the teacher is given a task based on the students' misconceptions. In their question, the choice for both students' reasoning is also based on the works of Cooper and Shore (2008), del-Mas and Liu (2005) and Meletiou-Mavrotheris and Lee (2005). The first student's response was designed to be influenced by the number of bars, which is not an indicator of a large standard deviation. By wrongly associating the group with the highest number of different values for the deviations, group C, with the distribution having the greatest standard deviation, the student excludes the deviations' sizes and the number of students associated with each deviation. Also, by following this logic, it would be difficult to identify the group with the smallest standard deviation because the other two groups (A and B) have the same number of bars. The second student is influenced by the non-variation in the height of bars in group C's distribution. By thinking this way, the student refers to the variation in the number of students instead of the variation in the quantity of water drunk by students monthly.

The cases above were aimed at observing how the interrogated teachers dealt with the students' conception of variability to better understand the type of interventions the teachers would choose. Depending on the teachers' answers, other questions were also asked during the interview in order to clarify their remarks and to obtain a deeper

²Adapted from Meletiou-Mavrotheris and Lee (2005).

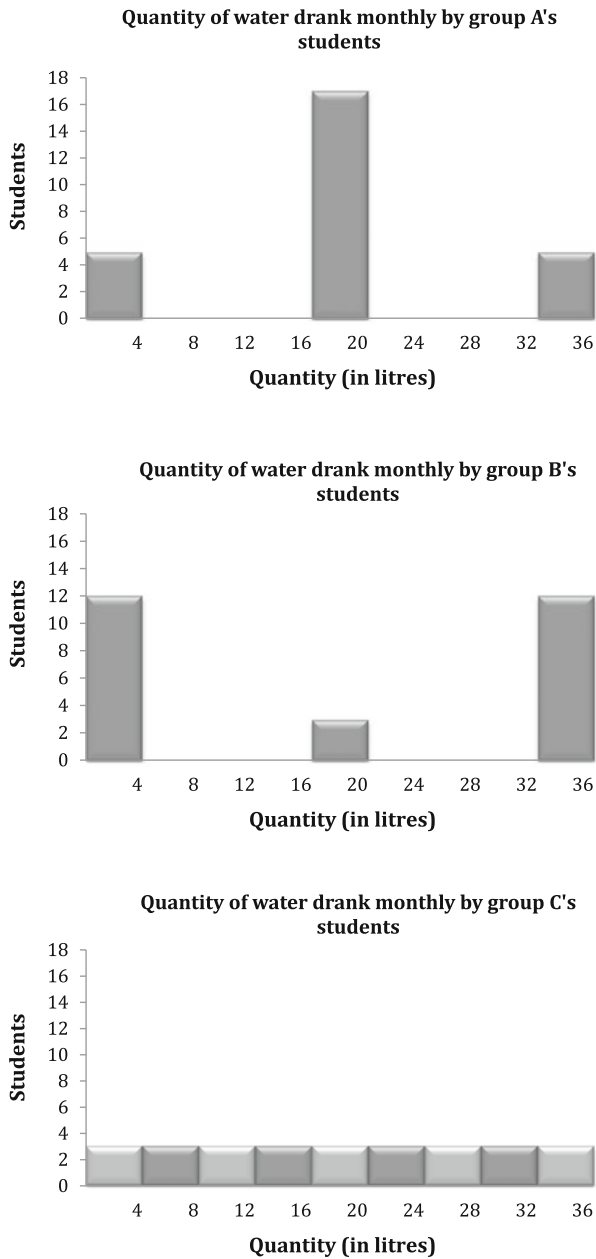


Fig. 10.4 A, B, C: quantity of water drunk monthly by secondary 4 students

comprehension of the professional knowledge of statistics the teachers used in relation to the concept of variability. No explanation was offered during the interview.

The interviews were taped, and the teachers and interviewers' comments were transcribed before being analyzed. The data were analyzed using the framework about the dimension of a statistical variable's values formulated in Sect. 10.2.1: assessing the responses as correct, incorrect, or unanswered; interpreting the alignment of responses with reasoning (conceptions). An inductive analysis process was favoured to identify categories from procedures identified by the researcher during the analytical process (Blais and Martineau 2006). It was possible to group the categories that emerged in the analysis under different themes. The main goal of the analysis presented here is to illustrate the nature of the strategies used by teachers in a statistical context involving the concept of variability.

10.4 Results

We will present the results coming from the two scenarios given individually to the twelve teachers. For each scenario, we present solutions provided by the teachers to the problem, followed by their interpretation of the students' solutions and a suggested intervention to make with students.

10.4.1 *First Case Example*

Solutions provided by the teachers

When asked to identify the graph illustrating the greatest variability in students' height, most teachers, i.e. 9 out of 12, identified school A's distribution as the one with the greatest variability. However, even though this is a good answer, the results must be examined in light of the arguments brought forward as all have different meanings. One teacher considered the number of bars. For this teacher, many bars equal a great variability and indicated a large variety of student heights.

Several different heights, so more dispersion and greater variability (translated from French).

This justification may refer to a meaning often found in the everyday language where large variety is associated with large variability. This common idea of variability differs from the statistical conception of the concept but still leads, in this case, to correctly identify school A as the one with the greatest variability in Secondary 1 students' height as its associated graph has more bars. At the same time, considering that both groups have the same number of students, a bigger number of bars may be associated with fewer students per height category. Even though this aspect was not suggested by the teacher in question, a comparison can be made with five other teachers' (of the group of nine teachers) arguments referring to the data concentration. These teachers explained their choice based on the density of the distribution of heights. They analyzed the data concentration by creating a visual example of the

data spread on the whole distribution. For example, school A's distribution has more different heights and fewer students for each height.

For school A, we see that few students are found in each height and that there is a bigger variety in height than in school B (translated from French).

Finally, the last three teachers who identified school A as the one with the greatest variability based their answer on the range. To answer the question, these teachers considered the spread between the distribution's minimum and maximum heights, which is 20 cm for school A's students and 14 cm for school B's. Using the range may be explained by the fact that this measure, which is simple to calculate, gives a quick first approximation of the data dispersion. However, its calculation is only based on the distribution's two extreme values. Used alone, the range is a limited method to measure variability. In this case, for example, this measure of dispersion doesn't consider the influence of frequency on variability for each bar. It is surprising to see the importance given to this measure of dispersion.

Among the three teachers who chose school B, two were influenced by the graphical representations' resemblance to a bell-shaped curve associated with a normal distribution. According to them, a large variability is associated with a distribution shape that deviates from the normal shape. The other teacher explained his choice based on the fact that school B's bars vary more in height. This teacher associates bar height with students' heights and not with frequency thus focusing on the variability of frequencies rather than on the variability of the variable in question, which is the height of the students.

Interpretation of the students' solutions and an intervention to make with students

In a second step, after analyzing the student response, five teachers out of 12 were not able to identify the mistake or at least recognize the students' misconceptions. For the three teachers among them who selected school B on their own resolution, it became difficult to refute the students' reasoning as the teachers not only came to the same final conclusions, but one also thought the same as the first student and the other two reasoned similarly to the second student in terms of the Gaussian curve being symmetrical. Each accepted the students' misconceptions using reasoning corresponding to their own. Two other teachers simply stated that they couldn't disagree as they were confused by the reasoning. These two teachers had both solved the problem by referring to the number of bars and to the range of each distribution of the heights.

The seven teachers who identified both students' misconceptions were able to suggest an intervention. Four teachers explained the problem by opposing the variability in sizes and frequencies to illustrate that in this case the problem needed to be solved horizontally and not vertically (the values corresponding to the mean are on the x -axis labeled with the different variable's values). The other three teachers who got the right solution favored an approach that allows the student to reflect on their misconceptions. For instance, one of them suggested tabulating the values so the shape of the distribution wouldn't influence the students. We named this approach

transition to numbers. The two other teachers gave the students a counter-example. For instance, one of them suggested a symmetrical distribution showing a low variability despite wide differences in the bars' heights.

If 14 students are 153 cm tall, 14 students 155 cm tall and 2 students 154 cm tall, you obtain high and low bars and the distribution is symmetrical. Do the student heights vary greatly? Not really, they all measure almost the same (translated from French).

For this task, seven teachers expressed knowledge of the conceptual issue by identifying the disruptive role of the graphic aspect. This knowledge was translated into explanations for students either by an alternate presentation of the problem (transition to numbers) or by giving them a counter example.

10.4.2 Second Case Example

Solutions provided by the teachers

Nine out of twelve teachers did not get the right solution. Of these nine teachers, one based his thinking on the data concentration of the central bin to solve the problem. By counting the number of people outside of the central bin, we can identify the distribution of groups B and C as those with the smallest standard deviation as they have the same amount of data outside of the central bin. This misconception doesn't include the value of the standard deviation of the data outside of the central bin.

For the large standard deviation, I selected B and C because 24 out of the 27 respondents were not in the average class (translated from French).

Another of these teachers was influenced by the symmetry of the three distributions. The teacher concluded that all of the distributions of the water drunk had the same standard deviation affirming that the positive deviations perfectly offset the negative ones. Here is what this teacher said:

I calculated the mean everywhere. I obtained 18. Then, I said that the distributions were symmetrical. Since the standard deviation is deviations's mean, I then affirmed that the three distributions had the same standard deviation [the teacher bases his thinking on the fact that the sum of the mean deviations is null for each distribution] (translated from French).

As for the other teachers who were incorrect, two of them specified that they couldn't answer the question. Three of them were influenced by the variation of the heights of the histograms' bars, just as the second student was, by claiming that the smallest standard deviation is in group C because of the uniformity of the bars height, and the greatest standard deviation is group A's because the graph shows a greater variation in the bars height.

The smallest was the third one because the data were more uniform [talking about the uniformity of the bar height of group C], and the biggest would be the first one because the data are more spread...the difference here between this and that [the subject shows the difference of the height of bars]. There is a bigger variation in the students, a difference of 12; from 17 to 5 (translated from French).

The other two teachers reasoned the same way as the first student by quantifying the number of possibilities (number of bars) for each group. For the biggest standard deviation, this reasoning leads to choosing group C as there are more possibilities in this group (more bars).

Less possibilities so less variability [speaking of the graphical representation of group A]. How many possible answers are there here? There are 1-2-3. This means that the standard deviation is smaller than group C's (translated from French).

Interpretation of the students' solutions and an intervention to make with students

As in the first case example, nine out of 12 teachers, including the five teachers that did not identify the students' misconception in the first case example, were not able to identify the issue or at least to see the student's misconceptions. It was difficult for them to intervene as they initially couldn't solve the problem, and several among them shared the student's misconceptions in their own resolution.

Of the seven teachers who disagreed with the students reasoning in the first case example, three of them disagreed with the reasoning of the two students in this case example and again, saw the issue of interpreting the data dispersion of the graphical representations in terms of data proximity to the center of the distribution. These teachers suggested an intervention that would have the students realize that Group C's graphical representation does not show the distribution with the greatest standard deviation nor with the smallest. For instance, one of the teachers explained this by referring to standard deviation measures.

I wouldn't know how to show him without the calculation [the subject refers to the standard deviation's measure for the three distributions]. There are obviously other methods, but I would be quite afraid to show another way, then in another situation where it wouldn't be done in such a way they would try to do it by reasoning and make mistakes. As a teacher, I often prefer to show them the so-called "safe" methods (translated from French).

The other two teachers suggested explanations by referring to the data concentration around the mean. For example:

To a student who says: "I think that Group C has the smallest dispersion because the bars all have the same height." I would answer negatively because extremes are still the same, i.e. between 0 and 4 and between 32 and 36, in graph A, there is a more even dispersion around the mean than in graph B where there is a cluster of people at the beginning, a cluster of people at the end and almost nobody in between. But with Group C, people are spread evenly (translated from French).

10.5 Discussion

This exploratory research project was based on the teachers' comprehension of the concept of variability in a statistical context through the exploration of two scenarios rooted in their practice, a concept that is at the heart of statistical thinking. The

proposed scenarios confronted teachers with students' answers and reasoning that highlighted misconceptions about variability and its graphical interpretation.

The knowledge of the concept and interpretation of variability, conceptions observed in school students and university students (Cooper and Shore 2008; Dabos 2011; delMas and Liu 2005; Meletiou-Mavrotheris and Lee 2005) are also observed in high school teachers. When asked to interpret the variability in a graphical representation of a distribution, some teachers were influenced by aspects associated with the distribution shape:

- Variability as a variation of the bars heights: The variation of the bar heights in a histogram becomes an indicator of the distribution's variability; the more the bars heights vary, the greater the variability. This is a misconception of variability.
- Variability as a deviation from normalcy: The variability of a distribution is determined by its resemblance or not to the Gaussian or normal curve; a low variability is associated with bell shaped "normal" distribution. This is a misconception of variability.
- Variability as an asymmetrical distribution: The variability of a distribution is determined by its symmetry, which in turn is associated with a low variability. This is a misconception of variability.
- Variability as a measure of variety: A distribution's variability is determined by the number of bars included in the graphical representation; a high number of bars represents a great variability. This conception refers to a meaning found in the common language where we associate a great variety with a great variability. This conception of variability, which ignores the size of the deviations and the number of students associated with each bar, is incorrect.

The resemblance between students and teacher's errors is important. It shows a phenomenon related to statistics that we must understand. Common conceptions of variability seem to interfere with the statistical notion of the concept. Both teachers and students mix everyday and statistical connotations. For example, it may be conceivable to associate uniformity with what varies little. This justification refers to a common language meaning and differs from the idea of the statistical concept. Furthermore, teachers' misconceptions show the influence of the shape of graphical representations of data. This created obstacles by encouraging misconceptions that would not have occurred otherwise, such as associating variability with the variation of bar heights. It also appears that the concept of standard deviation is meaningless for most respondents when its interpretation, measuring the variability in terms of the proximity of the data to the distribution's center, represents a challenge. These results support those of other research, which indicates that understanding the concept of standard deviation is often limited to applying its algorithm and consequently shows the difficulties in interpreting the results (Cooper and Shore 2010; Dabos 2011; delMas and Liu 2005).

The teachers who could identify the students' misconceptions stand among those who solved the problem correctly. The results show that teachers who were able to identify students' misconceptions could propose an intervention. In fact, the results linked to the two scenarios show a variety of interventions, each coming from very

different approaches. Whereas some teachers explicitly and systematically referred to the mathematical issues of the problem to help students clarify the concept and answer the question, others suggested directly ‘confronting’ students to make them doubt their reasoning. In the first case, no proposition was made to explicitly engage students in thinking, researching and validating their reasoning process that is based on the application of computational methods while in the second case teachers suggested conditions that would allow students to doubt their misconception and the answer obtained that way, thus creating a cognitive conflict forcing them to reevaluate their representations, to question their conception and to justify or change it if needed.

The results have methodological limitations due, for example, to the number of people questioned and limitations in the tasks given to the teachers, which involved only visual representations of data with no statistical measures (e.g., no information about standard deviation or mean). Nonetheless, the results do have some educational implications. First, these results might be useful for teaching students about variability; for instance, to connect informal notions of variability based on graphical displays to more formal measures of variability based on statistical indices related to dispersion such as standard deviation. Secondly for future teacher training, the variety of interventions offered by the teachers in the project could constitute the beginning of a reflection that could be used to train future teachers. Some interventions were found to be more creative providing conditions to allow students to become aware of their misconceptions. More importantly is the realization that some of the teachers could react instantly to students’ answers and reasoning while others could not. This context raises concerns and highlights the need to increase teacher training in statistics to expand their ability to intervene in the classroom in a statistical context and to develop students’ statistical thinking. Focusing on teachers’ professional knowledge in a statistical context and on how they use this knowledge in class is even more important because of the specificity of statistical reasoning. It is necessary to expand this knowledge, so it can be better understood and used by teachers at the appropriate moment in their practice. Research on students’ learning is necessary to serve as the basis for creating learning situations linked with a teaching/learning context that would allow teachers to become comfortable with how students reason in a statistical context. This would provide teachers with opportunities to learn how to intervene to improve students’ reasoning and statistical knowledge.

The results also bring considerations for future research about teaching variability. Because teaching the concept of variability seems to be a challenge for teachers, we think it is essential to continue research efforts to try and better understand the issues met by teachers in this study and consequently think of potential solutions to contribute to their professional development. We believe that more research is necessary to gain insight on teaching this concept in classes. These studies would help answer many questions unanswered in this text: How is the concept of variability currently taught by mathematics high school teachers? Do they encourage their students to have a global understanding of the concept (e.g., interpreting the spread of the data in terms of deviation from the distribution’s center, interpreting variability in a graphical representation, interpreting variability with dispersion measures, etc.)? What tasks do they use to teach the concept of variability? What context do

they use to make the tasks they give to students more meaningful when teaching this concept? What teaching resources do they use and what do they propose for teaching the concept of variability? What teaching strategies are the most successful? How can we help teachers learn to use those strategies that have shown promise?

We believe that such studies would help identify current teachers' needs when teaching the concept of variability, and contribute to the initial and continuous teacher training for teaching statistics in high school.

References

- Baillargeon, N. (2005). *Petit cours d'autodéfense intellectuelle (Short intellectual self-defense course)*. Montreal: Lux Publisher.
- 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 Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute.
- Bednarz, N., & Proulx, J. (2009). Connaissance et utilisation des mathématiques dans l'enseignement: Clarifications conceptuelles et épistémologiques (Knowledge and use of mathematics in teaching: Conceptual and epistemological clarifications). *For the Learning of Mathematics*, 29(3), 11–17.
- Bednarz, N., & Proulx, J. (2010). Processus de recherche-formation et développement professionnel des enseignants de mathématiques: Exploration de mathématiques enracinées dans leurs pratiques. (Research-training process and professional development of mathematics teachers: Exploration of mathematics rooted in their practice). *Éducation et Formation (Education and Training)*, 293, 21–36.
- Blais, M., & Martineau, S. (2006). L'analyse inductive générale: Description d'une démarche visant à donner un sens à des données brutes. (General inductive analysis: Description of a process aiming at giving meaning to raw data). *Recherches qualitatives (Qualitative Researches)*, 26(2), 1–18.
- Bloch, I. (2009). Les interactions mathématiques entre professeurs et élèves. Comment travailler leur pertinence en formation? (Mathematical interactions between teachers and students. How to make them relevant in training?). *Petit x*, 81, 25–52.
- Borim da Sina, C., & Coutinho, C. (2008). 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*.
- Brousseau, G. (1998). *Théorie des situations didactiques (Theory of didactical situations)*. Paris: La pensée sauvage Publishers.
- Canada, D. (2004). *Elementary preservice teachers' conceptions of variation* (Doctoral dissertation). Portland State University, Portland, OR.
- Canada, D. (2006). Elementary pre-service teachers' conceptions of variation in a probability context. *Statistics Education Research Journal*, 5(1), 36–63.
- Cooper, L., & Shore, F. (2008). Students' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. *Journal of Statistics Education*, 15(2), 1–13.
- Cooper, L., & Shore, F. (2010). The effects of data and graph type on concepts and visualizations of variability. *Journal of Statistics Education*, 18(2), 1–16.
- Dabos, M. (2011). *Two-year college mathematics instructors' conceptions of variation* (Doctorate in education thesis). University of California, Santa Barbara, CA.

- Davis, B., & Simmt, E. (2006). Mathematics-for-teaching: An ongoing investigation of the mathematics that teachers (need to) know. *Educational Studies in Mathematics*, 61(3), 293–319.
- delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4(1), 55–82.
- Dodge, Y. (1993). *Statistics: encyclopedic dictionary*. Switzerland: Université de Neuchâtel.
- Even, R. (1993). Subject-matter knowledge and pedagogical content knowledge: Prospective secondary teachers and the function concept. *Journal for Research in Mathematics Education*, 24(2), 94–116.
- Even, R., & Tirosh, D. (1995). Subject-matter knowledge and knowledge about students as sources of teacher presentations of the subject-matter. *Educational Studies in Mathematics*, 29(1), 1–20.
- 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., delMas, R., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 117–147). New York, NY: Lawrence Erlbaum Associates.
- Gattuso, L., & Vermette, S. (2013). L'enseignement de statistique et probabilités au Canada et en Italie (The teaching of statistics and probability in Canada and Italy). *Statistique et Enseignement*, 4(1), 107–129.
- 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 Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, Arizona, USA*. Voorburg, The Netherlands: International Statistical Institute and International Association for Statistical Education.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.
- Holm, J., & Kajander, A. (2012). Interconnections of knowledge and beliefs in teaching mathematics. *Canadian Journal of Science, Mathematics and Technology Education*, 12(1), 7–21.
- Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. E. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.
- Margolinas, C. (2014). Concepts didactiques et perspectives sociologiques? (Didactical concepts and sociological perspectives?). *Revue Française de Pédagogie*, 188, 13–22.
- Meletiou-Mavrotheris, M., & Lee, C. (2005). Exploring introductory statistics students' understanding of variation in histograms. In *Proceedings of the 4th Congress of the European Society for Research in Mathematics Education, Sant Feliu de Guíxols, Spain*.
- Ministry of Education and Higher Education. (2004). Quebec education program (QEP) secondary: Cycle one. In *Mathematics*. Quebec: Government of Quebec.
- Ministry of Education and Higher Education. (2007). Quebec education program (QEP), secondary: Cycle two. In *Mathematics*. Quebec: Government of Quebec.
- Moreira, P., & David, M. (2005). *Mathematics in teacher education versus mathematics in teaching practice: A revealing confrontation*. Paper presented at the conference of the 15th ICMI study on the Professional Education and Development of Teachers of Mathematics, Águas de Lindóia, Brazil.
- Moreira, P., & David, M. (2008). Academic mathematics and mathematical knowledge needed in school teaching practice: Some conflicting elements. *Journal for Mathematics Teacher Education*, 11(1), 23–40.
- 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.
- Proulx, J. (2008). Exploring school mathematics as a source for pedagogic reflections in teacher education. *Canadian Journal of Science, Mathematics and Technology Education*, 8(4), 331–354.

- Proulx, J. & Bednarz, N. (2010). Formation mathématique des enseignants du secondaire. Partie 1: Réflexions fondées sur une analyse des recherches (High school mathematics teacher training. Part 1: Reflexions based on a research analysis). *Revista de Educação Matemática e Tecnológica Ibero-Americana*, 1(1). <http://emteia.gente.eti.br/index.php/emteia>.
- Proulx, J. & Bednarz, N. (2011). Formation mathématique des enseignants du secondaire. Partie 2: Une entrée potentielle par les mathématiques professionnelles de l'enseignant (High school mathematics teacher training. Part 2: A potential entry by the teacher's professional mathematics). *Revista de Educação Matemática e Tecnológica Ibero-Americana*, 1(2). <http://emteia.gente.eti.br/index.php/emteia>.
- 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: Kluwer Academic Publishers.
- Sanchez, E., Borim da Sina, 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, Germany: Springer.
- Savard, A. (2014). Developing probabilistic thinking: What about people's conceptions? In E. Chernoff & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (Vol. 2, pp. 283–298). Berlin: Springer.
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. (1988). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Whittrock (Ed.), *Handbook of research on teaching* (pp. 3–35). New York, NY: Macmillan Publishers.
- Silva, C. B., & Coutinho, C. Q. S. (2006). The variation concept: A study with secondary school mathematics teachers. In A. Rossman & B. Chance (Eds), *Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg: International Statistical Institute and International Association for Statistical Education.
- Vergne, C. (2004). La notion de variabilité dans les programmes de seconde (2000)—Étude de conditions de viabilité didactique (The concept of variability in secondary programs (2000)—A study of conditions for the viability of didactics). In *Actes des XXXVI^{èmes} journées de Statistique, Société Française de Statistique (Acts from the XXXVIst days of Statistics, French Society for Statistics)*, Montpellier, France.
- Wozniak, F. (2005). *Conditions et contraintes de l'enseignement de la statistique en classe de seconde générale. Un repérage didactique (Conditions and constraints of teaching statistics in general secondary classes. A didactical identification)* (Doctoral dissertation). Université Claude Bernard Lyon 1, Lyon.

Chapter 11

Secondary Teachers' Learning: Measures of Variation



Susan A. Peters and Amy Stokes-Levine

Abstract This chapter describes results from a project to design and implement professional development for middle and high school mathematics teachers to investigate how dilemma, critical reflection, and rational discourse affect teachers' understandings and reasoning about variation. Framed by transformative learning theory, this study highlights how teachers' engagement with activities designed to prompt dilemma, consideration of multiple perspectives through multiple representations and rational discourse, and examination of premises underlying measures and procedures broadened teachers' perspectives on measures of variation. This study contributes to teacher education by identifying circumstances conducive to deepening statistical understandings and supporting increasingly complex statistical reasoning.

Keywords Mean absolute deviation · Professional development
Standard deviation · Transformative learning theory · Variation

11.1 Background

Statisticians and statistics educators espouse the importance of understanding variability for statistical thinking (e.g. Wild and Pfannkuch 1999). Even though the primacy of variation to statistics has long been accepted, researchers are only beginning to uncover the complexities of developing conceptual understandings of variability. Considerable research focuses on students' intuitive notions of variability (e.g. Reading and Shaughnessy 2004; Watson 2006) and school students' and adults' limited understandings of variability and formal measures of variation (e.g. deLMas and Liu 2005). Few studies focus on teachers, particularly in-service teachers, and how to facilitate their development of conceptual understandings of variability (Sánchez

S. A. Peters (✉) · A. Stokes-Levine
University of Louisville, Louisville, KY, USA
e-mail: s.peters@louisville.edu

A. Stokes-Levine
e-mail: amy.stokes-levine@jefferson.kyschools.us

et al. 2011). Yet, teachers need deep understandings in order to facilitate student learning. Continued research is needed to design instruction and activities that are effective for teachers to develop conceptual understandings as one step towards achieving larger educational goals. We describe an investigation of a professional development program designed to support middle and high school mathematics teachers' reasoning and learning in statistics. We focus on teachers' engagement with activities to address how dilemma, critical reflection, and rational discourse affect teachers' understandings and reasoning about measures of variation.

11.2 Previous Research About Measures of Variation

Although school and tertiary students in general may have less sophisticated mathematical understandings than secondary mathematics teachers, Shaughnessy (2007) suggests that many "teachers have the same difficulties with statistical concepts as the students they teach" (p. 1000)—a view that is supported by reviews examining students' and teachers' reasoning and understanding of variation (Sánchez et al. 2011). As a result, the body of research that examines students' reasoning, understanding, and learning about variation and related concepts can provide insights into struggles that teachers might experience and important elements and connections needed for teachers to develop robust understandings of variation.

Research suggests that school students have intuitive conceptions about variation and are able to reason about the range and spread of data relative to a center (e.g. Reading and Shaughnessy 2004). Additionally, students exhibit improved reasoning about variability throughout their educational years as they study ideas related to data and chance (e.g. Watson et al. 2007; Watson and Kelly 2004a, b, 2005). In general, however, students encounter difficulties when asked to reason about variation using formal measures of variation.

Developing meaning for and reasoning about formal measures of variation, particularly standard deviation, appears to be particularly problematic for students. For example, prior to beginning an introductory-level college statistics course, some mathematics majors struggled to describe the meaning of standard deviation beyond relating it to variability in data (Cook and Fukawa-Connelly 2016). When comparing distributions using formal measures of variation, other first-year college students tended to examine agreement among different measures of variation to compare variability in distributions in place of choosing measures for comparison based upon data characteristics (Lann and Falk 2003). Their comparisons typically relied on the range more than standard deviation/variance, mean absolute deviation, or interquartile range. To draw conclusions about the magnitude of standard deviation in two distributions, some introductory-level statistics students compared standard deviations for multiple pairs of distributions by attempting to create rules to generalize patterns in histogram bars (delMas and Liu 2005). Very few students employed conceptual approaches to coordinate the location of the mean with deviations from the mean. Even students who successfully completed an introductory statistics course

with top grades viewed standard deviation only as a rule or formula (Mathews and Clark 2003). In general, students who complete introductory courses—even courses designed to advance reasoning about variation with formal measures of variation—may only begin to consider variation as a measure of spread from center and begin to display advanced understandings of variation by applying their knowledge to novel situations (Garfield et al. 2007).

Teachers experience some of the same difficulties that tertiary students exhibit when reasoning about and with formal measures of variation. For example, few of the prospective science and mathematics teachers participating in Makar and Confrey's (2005) study compared data sets by using standard deviation. Makar and Confrey observed that standard deviation as a measure of variation held little meaning for these teachers. Little evidence suggests that many preservice secondary mathematics teachers understand formal measures of variation as anything more than numerical values or as computations (e.g. Makar and Confrey 2005; Sorto 2004). For many teachers, even if they are able to calculate values for standard deviation and discuss standard deviation as a measure of variation, they may be unable to reason about standard deviation in conjunction with the mean (Clark et al. 2007; Silva and Coutinho 2008).

Relatively recent interventions with students, however, offer strategies for helping students (and teachers) to overcome some of these struggles associated with developing understanding and reasoning about measures of variation for students to successfully reason about and with representations and measures such as determining which best represent data (Garfield and Ben-Zvi 2005) and reasoning about spread relative to center (Garfield et al. 2007). By engaging elementary students in processes of modeling data that included inventing and revising measures of variability for data from a measurement context, students invented and explored measures that included sum of deviations and sum of absolute deviations from the mean and median (Lehrer and Kim 2009; Lehrer et al. 2007), average absolute deviation from the mean and median (Lehrer et al. 2007), and interquartile range (Lehrer et al. 2007). Such explorations set the stage for students to engage meaningfully with conventional measures. Work with tertiary students suggests that an approach in which students consider the constraints and affordances of different formal measures of variation can be effective for deepening students' conceptual understandings of the measures (Garfield et al. 2007). These studies offer insights into the types of activities that might facilitate teachers' development of conceptual understandings of formal measures of variation. Further insights can be gleaned from theories of adult learning.

11.3 Frameworks

Most Pre-K–12 instruction focuses students on answering questions of what and how, which allows students to construct understandings of new ideas or to enhance their current understandings. Adult learning, however, often results from focusing on the premises behind content and processes towards answering questions of why.

Transformative learning theory (Mezirow 1991) is an adult learning theory that takes into account the transformative learning that results from reflecting on premises. An overarching tenet of the theory is that powerful learning results from transforming meaning perspectives, which are the broad predispositions formed from culture and experiences (Cranton 2006; Mezirow 1991).

Perspective transformation can begin with events that trigger a “disorienting dilemma” to prompt examination of broad presuppositions or with a series of incremental dilemmas that prompt examination of particular knowledge or attitudes (Taylor 2000). These dilemmas are feelings of dissatisfaction with current thinking or knowledge in relation to meaning perspectives or meaning schemes, which consist of specific expectations and knowledge used to interpret experiences implicitly (Cranton 2006; Mezirow 1991). Dilemmas that induce questioning of assumptions for meaning schemes can be resolved by creating, enhancing, or transforming meaning schemes (Mezirow 2000). For example, a teacher whose initial meaning scheme for standard deviation consists of algorithmic steps for calculating a value might experience dilemma when engaging with activities to develop understanding of the standard deviation as the approximate average deviation from the mean and subsequently enhance her meaning scheme by viewing the formula in terms of average and deviations. The teacher might transform her meaning scheme for variation by rejecting her prior computational conceptions of measures of variation to view variation as a multifaceted construct that permeates all of statistics after studying experimental design and reflecting on sources of variability such as measurement error. Critical reflection—reflecting on premises to question the importance, validity, or utility of knowledge—is crucial for transformation to occur. Critical reflection often is supported by rational discourse—dialogue with oneself or others to examine alternative perspectives and to assess expectations and knowledge—towards developing and acting on plans to resolve dilemmas.

Transformative learning theory framed prior retrospective research that identified factors teachers perceived as contributing to their development of deep understandings of variation. Important for teachers’ learning was not only engagement in rational discourse to consider the perspectives of others but also active exploration with data using multiple representations to gain new perspectives for concepts such as sampling distribution, to explore premises underlying concepts, and to justify methods and conclusions (Peters and Kopeikin 2016). Teachers also described how actively engaging with tasks and activities that were focused on fundamental statistical concepts and principles and key aspects of variability were important for their learning about variation and related concepts (Peters 2014; Peters and Kopeikin 2016).

We designed our professional development program to build on the results of this research and research related to students’ reasoning and learning about statistical variation. We designed the professional development activities to include planned triggers for dilemmas in typical areas of struggle using an overarching framework for reasoning about variation (Peters 2011). The program offered teachers considerable opportunities to explore content conceptually, engage in rational discourse with other teachers, and examine underlying premises and reflect critically on the content to enhance or transform their skills and knowledge related to formal measures of vari-

ation. The program also incorporated characteristics of “high quality” professional development such as sustained duration and focus on content towards developing teachers’ content knowledge (e.g. Goos et al. 2007; Smith et al. 2005). The program included a one-week intensive summer experience in which teachers actively experienced K–12 statistics content as learners using the problem solving process detailed in the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE; Franklin et al. 2007) and focusing on variation throughout the process.

Activities to investigate measures of variation began with teachers comparing average five-year rates of return on stocks, mutual funds, and certificates of deposit using a variety of representations generated in TinkerPlots (Konold and Miller 2005) and considering students’ reasoning associated with boxplots and interquartile range to describe and compare distributions by examining variability within distributions in addition to overlap and variability between distributions. Teachers’ explorations continued by using arm circumference measurements that they collected and that were collected from the National Health and Nutrition Examination Survey to consider distributional features captured and not captured by a variety of standard and non-standard measures for variation. Originating from research on students’ reasoning about variation, these measures included the range (e.g. Garfield et al. 2007), sum of deviations and sum of absolute deviations from the mean and median (e.g. Lehrer and Kim 2009; Lehrer et al. 2007), average absolute deviation from the mean and median (e.g. Lehrer et al. 2007), and interquartile range (Biehler 2007; Wild et al. 2011). Teachers then examined mean absolute deviation and its properties, contrasted mean absolute deviation with standard deviation both symbolically and visually through multiple representations, considered why standard deviation typically is used in place of mean absolute deviation, and considered the effects of outliers on each measure to make progress towards developing dynamic conceptions of mean absolute deviation and standard deviation by coordinating changes to the relative values about a mean with their deviations from the mean. Lastly, teachers explored standard deviation further by playing the standard deviation game developed by delMas and described in delMas and Liu (2005).

11.4 Data Sources and Methods

Ten middle- and nine high-school teachers participated in a one-week, 40-hour, summer professional development program. Teachers varied in their statistical learning and teaching experiences. Every teacher completed a minimum of one introductory-level statistics course as part of a secondary, undergraduate, or graduate program, and four teachers completed one or more advanced or mathematical statistics courses. During a typical school year, some teachers taught at most 10 statistics-related lessons, whereas others taught the equivalent of an introductory college-level statistics course.

Data sources included audio- and video-recordings of large-group discussions and small-group activities from professional development sessions, teachers’ written

work and reflections, one semi-structured interview with each teacher, and teachers' results on pre- and post-tests developed by the Levels of Conceptual Understanding in Statistics (LOCUS) project (Jacobbe et al. 2011). LOCUS forms used for pre- and post-assessment are equated and consist of 50 multiple-choice questions that cover the gamut of statistical problem solving phases (i.e., formulate questions, collect data, analyze data, interpret results) across beginning, intermediate, and advanced statistical literacy levels.¹ Semi-structured interviews were conducted after the weeklong summer professional development sessions concluded to provide insights into how dilemma, critical reflection, and rational discourse affected teachers' understandings and reasoning. Teachers also responded to statistics problems with multiple entry points for the researchers to examine their statistical understandings and reasoning. We gained further insight into teachers' understandings and reasoning from professional development session recordings and teachers' written work.

For each recording, we created a log of the content and developed transcripts. We first examined aspects of the professional development program designed to encourage dilemma, critical reflection, and rational discourse by identifying relevant transcript passages. We searched written work and transcripts for evidence of dilemma, critical reflection, and rational discourse, paying attention to indications of insights, questions, or confusion; thoughts and reasoning beyond the immediately observable; content-related interactions with other teachers, students, or texts or consideration of multiple perspectives; and references to the preceding as potential evidence for dilemma, critical reflection, and rational discourse, respectively. Two researchers separately analyzed each transcript and teachers' work and reflections using a combination of codes developed from the theoretical framework and codes that emerged from the data. Theory-related codes included the elements of transformative learning (e.g. disorienting dilemma, self-examination, assessment of assumptions, engaging in rational discourse, etc.) identified by Mezirow (1991, 2000). We examined data for each participant, discussed discrepancies in our analyses until we reached agreement, and made comparisons across participants to look for common themes as well as variations from the themes.

11.5 Results

From the beginning to the end of the week of the summer professional development program, teachers' mean scores on equated LOCUS assessment forms improved significantly from 75.5 to 81.1% ($t(18) = 3.53, p = 0.001$), and their median scores improved significantly from 76 to 80% ($W = 137.5, p = 0.002$). Breaking results apart by grade-level certification revealed that middle school teachers' mean scores improved significantly from 72.4 to 78.2% ($t(9) = 2.82, p = 0.010$) and median scores improved significantly from 71 to 78% ($W = 34.5, p = 0.013$), and high school

¹The beginning, intermediate, and advanced statistical literacy levels assessed by LOCUS correspond with the A, B, and C levels of development, respectively, articulated in the GAISE report (Franklin et al. 2007).

teachers' mean scores improved significantly from 78.89 to 84.22% ($t(8) = 2.08$, $p = 0.035$) and median scores improved significantly from 78 to 86% ($W = 37.5$, $p = 0.043$). Three of the 19 teachers scored lower on the posttest, answering one, two, or three more questions incorrectly on the posttest than on the pretest. The greatest improvement came from two teachers who answered nine and ten more questions correctly on the posttest.

Although this group of teachers did not show significant improvement on individual variation items, they did, on a scale from 1 to 10, report significantly increased knowledge for reasoning about interquartile range (mean increase of 1.63, $t(18) = 4.633$, $p < 0.001$), calculating mean absolute deviation and standard deviation (mean increase of 3.34, $t(18) = 6.909$, $p < 0.001$), and reasoning about mean absolute deviation and standard deviation (mean increase of 2.79, $t(18) = 7.512$, $p < 0.001$). In their final reflections on the week, eight teachers explicitly mentioned deepened understandings of mean absolute deviation and standard deviation as important learning they experienced from professional development activities. Preliminary analyses suggest several features that may have contributed to teachers' overall improved scores and perceptions of increased knowledge and reasoning abilities.

11.5.1 Dilemma

Teachers experienced dilemmas throughout the professional development program. Several dilemmas stemmed from interpreting nonstandard measures of variation and considering distributional features captured by the measures. For example, after examining the sum of absolute deviations from the mean, Bob² struggled to evaluate distributional features captured by the sum, noting, "I don't know what that really does tell us." When comparing the sum of absolute deviations with the mean absolute deviation, Naomi queried, "I wonder what the benefit is of using one over the other." Although many of the teachers previously encountered ideas of deviation, absolute deviation, and mean absolute deviation, few, if any, previously considered nonstandard measures such as the sum of absolute deviations from the mean. Focusing on why they might use a nonstandard measure and considering underlying premises such as why mean absolute deviation might better describe distributions than other measures caused teachers to reexamine their understandings.

Other dilemmas arose when teachers interpreted graphical displays of data that differed from traditional displays of univariate data such as dotplots or histograms. For example, when interpreting graphs displaying linear deviations of upper arm circumferences in centimeters from a mean upper arm circumference of 32.6 cm (Fig. 11.1), teachers struggled to coordinate the graph with their procedures for calculating mean absolute deviation. For example, the group consisting of Jackson, Mallory, and Daphne calculated a mean absolute deviation of 2.225 cm for the data displayed in Fig. 11.1 and observed that the mean absolute deviation was larger than

²All teacher names used in this chapter are pseudonyms.

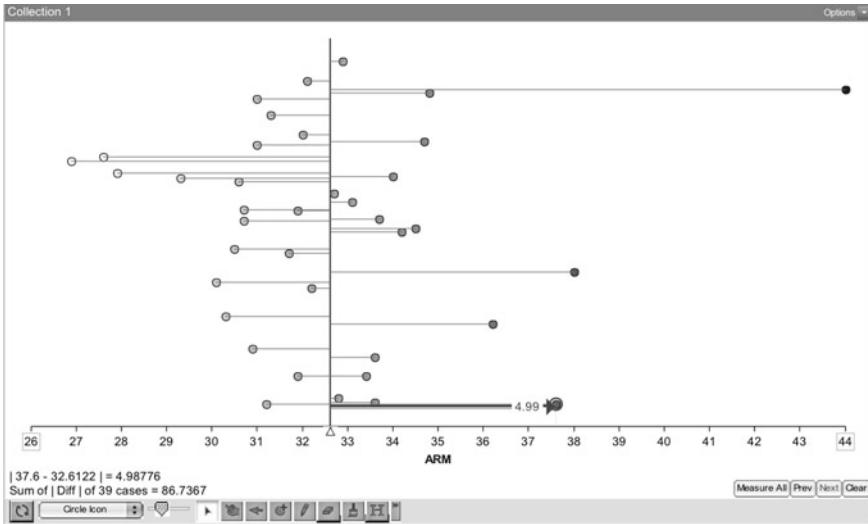


Fig. 11.1 Deviations from a mean upper arm circumference of 32.6 cm

the majority of absolute deviations due to the large data value of 44 cm. However, they seemingly struggled to reconcile the differences between the averages they calculated and between distance and deviation.

Jackson: The line is the, the vertical line is the, is the mean. And then they've got a line from each dot to the mean.

Mallory: Well the mean [*MAD*] that we calculated was 2.225. This [*points to vertical line*] is [not 2.225].

Jackson: No, no, no. Okay, we're getting our means [mixed up]. Both are means. So the 2.225 is the mean of the deviations from the mean.

...

Mallory: Oh!

Jackson: This [*points to vertical line on graph*] is the mean of the actual arms.

Mallory: Okay.

...

Daphne: What do the horizontal lines represent?

Mallory: The distance from the mean.

Jackson: The deviations.

Daphne: Same thing. Right? ... Now the question is, how does the mean of the deviations you calculated relate to the graph below. Means the average of all those.

Jackson: So the average length of the horizontal lines is 2.225.

Daphne: Is the average length of the lines, right? That's what you just said. The mean of the deviations.

Mallory seemingly became confused after calculating a mean absolute deviation value and hearing Jackson refer to the vertical line at 32.6 cm as the mean. After Jackson clarifies the two means under consideration in the problem, the three teachers consider what the horizontal segments represent, with Daphne equating Mallory's description of the representation of horizontal line segments as distance from the mean with Jackson's description of the representation as deviations. Ultimately, the group was able to make connections between the calculation for mean absolute deviation and the graphical display. Other groups struggled to make connections because they focused on deviation rather than distance, such as Landen who concluded, "so the mean of the deviation that we calculated doesn't relate to this graph because we calculated the mean of the absolute deviations." As teachers responsible for teaching mean absolute deviation (or standard deviation), Mallory and other teachers knew how to calculate values for the measure(s) they taught. However, the graph presented a novel perspective for considering mean absolute deviation—one focused on highlighting defining properties of the measures and the meaning of the measures—that raised questions for the teachers.

Other teachers experienced dilemmas as they struggled to connect newly encountered ideas with concepts they thought they understood. For example, Bryce's understanding of standard deviation was perturbed when he encountered mean absolute deviation: "When I see average deviation I think standard deviation. So...average deviation must mean some other measurement I'm not really familiar with." As a high school teacher of an introductory statistics course, Bryce had not previously encountered mean absolute deviation. Recognizing the existence of a measure different from standard deviation that aligned better with his definition for standard deviation provoked dilemma.

Some of the teachers' dilemmas came from planned triggers—questions included in activities to provoke dilemmas that could be anticipated. These dilemmas included dilemmas stemming from nonstandard measures and graphical displays. Other dilemmas were unanticipated such as high school teachers confusing standard deviation with mean absolute deviation. Whether teachers encountered anticipated or unanticipated dilemmas, however, they sought resolution to their dilemmas, often by enlisting the help of other teachers.

11.5.2 Rational Discourse and Alternative Perspectives

Teachers engaged in rational discourse with others and examined multiple representations to gain alternative perspectives while considering premises, yielding insights to resolve dilemmas. Sally, for example, indicated that she learned more about mean absolute deviation as a result of: "a lot of the group work that we did and just hearing other people explain it to me, other than like reading it in a book or online,

understanding what it means a little bit more.” Similar to Sally, a majority of the teachers identified working through activities as students and in collaboration with other teachers as beneficial for their learning, and some of the richest interactions occurred between middle and high school teachers due to the different content that they taught.³ Consider the interactions among, for example, Bob, Landen, and Audrey as they examined the sum of absolute deviations from the mean.

Bob: Sum of the absolute deviations from the mean. I don’t know what that really does tell me. It’s just...

Landen: What I thought is it tells you if it, if there’s a whole bunch of deviation from the mean.

Bob: Yeah. If the numbers are huge.

Landen: Yeah, because if the number’s really big then...

Audrey: Because that’s really, that’s basically mean absolute, that’s the MAD [mean absolute deviation] that we talk about. And that tells you...

Bob: But that’s before you do the MAD so you have the answer before you do your MAD.

Audrey: Oh, yeah!

Landen: So it tells you if there’s a whole bunch. If, if there’s some that are really, really big... Cause obviously, the smaller the number that it is.

Audrey: But it’s very relative too. You have to know what you’re starting with. I don’t know.

Landen: It’s relative to the size and the numbers in the data set.

Audrey: Yes. Yes.

Landen: But I mean it’s really a measure of preciseness.

Bob: Measure of spread.

As middle school teachers who taught mean absolute deviation, Bob and Audrey contributed observations of similarities and differences between mean absolute deviation and the sum of absolute deviations from the mean to the discussion. Landen, a high school teacher who previously “did not know MAD,” focused on aligning their observations with their calculations for the three teachers to reach consensus that the sum provided a measure of spread or preciseness that should be considered in relation to the number of and magnitude of data in the data set. Each teacher contributed different features to advance each teacher’s understandings of the measure.

In addition to group discussions providing multiple perspectives of content, teachers credited multiple representations with providing alternative perspectives that served to enhance their understandings. For example, teachers such as Daphne identified explorations with Fathom (Finzer 2002) and graphs such as those displayed in Fig. 11.1 and in Fig. 11.2⁴ with enabling them to visualize differences between

³In the United States, the Common Core State Standards in Mathematics (National Governors Association Center for Best Practices and Council of Chief State School Officers 2010) include mean absolute deviation in the standards for sixth grade; in high school, the focus shifts from using mean absolute deviation as a measure of variation to using standard deviation.

⁴In Fathom, when selecting the option to Show Squares on a bivariate graph, the software presents a visual representation of a square using the vertical segment between the data point and the fit

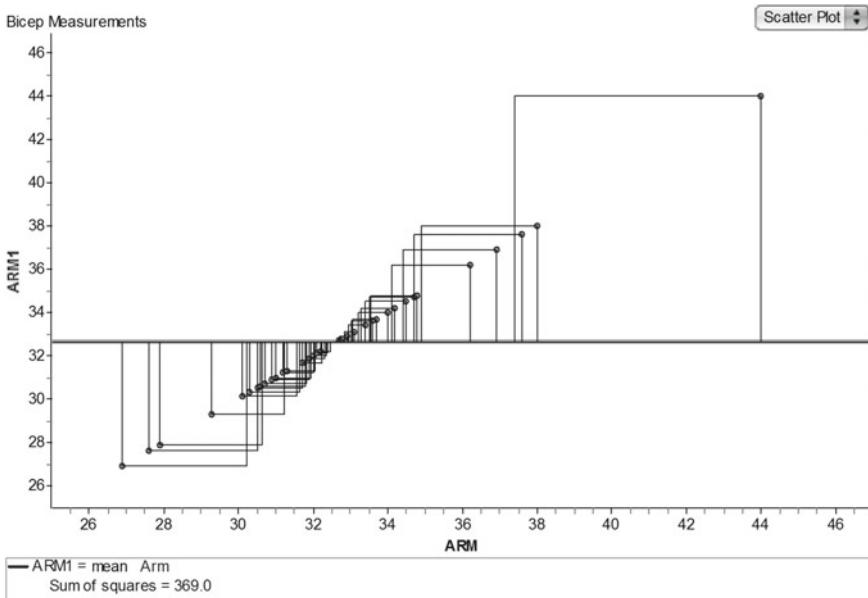


Fig. 11.2 Fathom representation of squared deviations from a mean of 32.6 cm for a univariate data set of upper arm circumferences in centimeters [ARM and ARM1]

mean absolute deviation and standard deviation such as how outliers influence each measure in ways that the symbolic formulas for calculating the measures could not. Specifically, Daphne identified working with “different data sets on Fathom with and without outliers and seeing the changes in values” as more effective for her to consider the effects of outliers on measures than calculating values for the measures without any type of visualization. In particular, she mentioned the importance of seeing the area representation for squared deviations in comparison with the length representation for absolute deviations. Brittany identified graphical explorations with Fathom for helping her to understand differences between mean absolute deviation and standard deviation: “I had never heard of MAD before this week, but how to calculate and why makes sense—I have only calculated SD [standard deviation] by hand a few times due to the ease of a graphing calculator but building it through Fathom helped me to understand.” She further noted how “the visuals provided by the graphs helped to make comparisons,” which were further facilitated through recording similarities and differences in “the matrices/charts [that] helped pull it altogether.”

In many cases, individual representations from student work or from TinkerPlots were representations teachers had not previously encountered, such as case-value

line as the basis for the square. The software does not use the side lengths to form squares with congruent sides.

plots,⁵ hat plots,⁶ or hat plots and boxplots superimposed over dot plots, which caused them to also question the utility of the representations. In other cases, using multiple representations to represent data allowed teachers such as Daphne to consider “what different representations show you and don’t show you” about distributions and to get what Rachel called “a clearer picture of variability.” The teachers used the representations to begin considering premises such as when Daphne considered circumstances under which “different representations are ‘better’” for describing data or comparing distributions. Similarly, Jessica noted the utility of using dynamic software to generate graphical representations “to see the comparisons and get an understanding of why” one representation might be better than another.

Beyond using multiple representations to explore premises, teachers engaged in rational discourse to examine premises. For example, the following excerpt reveals how teachers examined the effect of outliers on standard deviation in contrast with mean absolute deviation.

- Cecilia: How would this [squaring large deviations] affect *the mean* of the squared deviations?
- Lewis: Well depending on, depending on outliers it’s going, er, not even outliers but like values that are far away from the mean, it’s going to have a *huge* effect...
- Naomi: So values farther from the mean will be, um—like will have more of an effect?...the sum of squares...
- Cecilia: Greater impact when squared than absolute value...
- Lewis: But why? [*Pause.*] I think what we’re trying to get at here is that the more varied our data is, the more this value [mean of squared deviations] will be.
- Naomi: So just basically, the larger the deviation the larger the value [mean of squared deviations] will be.

While examining graphs such as those displayed in Figs. 11.1 and Fig. 11.2, focusing on sums and means of absolute and squared deviations, and collaboratively contributing to conversations to answer questions, Cecilia, Lewis, and Naomi were able to conclude that large outliers and values far from the mean would have a greater effect on standard deviation than on mean absolute deviation. As with most of the rational discourse the teachers evidenced during the professional development program, teachers resolved minor dilemmas related to statistics, such as Naomi’s questioning the effect of values at a large distance from the mean, by interacting with other teachers and focusing on premise-related questions. In some cases, however,

⁵A case-value plot is a bar graph representation in which each data value is represented by a bar with a length that represents the magnitude of the data value.

⁶A hat plot is a representation of data that is similar to a boxplot in that it represents a middle collection of data values in a data set using a rectangle and extends the bottom edge of the rectangle to the minimum and maximum data values to produce a visual that resembles a hat. Unlike a boxplot, the central box does not necessarily represent the middle 50% of data but may instead represent the middle 1/3 of data or the data within one standard deviation of the mean.

rational discourse occurred after teachers previously engaged in critical reflection to consider premises.

11.5.3 Critical Reflection

The professional development activities were designed to provoke both reflection on content and processes and critical reflection on premises by constantly focusing teachers on examining statistical concepts and understanding the premises behind statistical techniques. Approximately half of the teachers credited this focus on answering questions of “why” with deepening their understandings of the content and contrasted this focus with prior experiences. Jackson acknowledged the unique (to him) focus on premises when he discussed his experiences in three previous statistics courses.

I've taken three stats classes. I've taken Psych[ology] stats and two different stats classes. We talked about standard deviation in every single one of them. The amount of time that I've spent today attempting to actually understand [standard deviation] is greater than the amount of time that I spent attempting to understand in all of the other classes combined.

In anonymous feedback on the week of professional development activities, several teachers noted the benefit of finally understanding the concepts explored during the professional development program, indicating that they finally knew “what standard deviation really is” or “what mean absolute deviation really is...this confused me my first year!” Although teachers can be heard commenting throughout the week's videos about how the professional development activities pushed their thinking, such as Bob indicating that his “mind was just blown” or Mallory indicating that her brain was “squishy” when investigating the conceptual underpinnings of mean absolute deviation and standard deviation, most indicated benefit from focusing on conceptual understanding and premises.

Several teachers attended the professional development program after previously spending significant time studying statistics and reflecting on the content during their preparations to teach the content. For example, Margaret taught a college-equivalent introductory statistics course and previously tried to determine why standard deviation typically is used in place of mean absolute deviation. She observed that the question about “why do we use this one?” arose “every year” in her classes. She also shared the following within her small group of teachers working together on activities:

For some reason, the standard deviation is a lot more useful than the mean absolute deviation...for some reason statisticians prefer this one...But there is a reason. I can't tell you what it is but it does exist. Why they use standard deviation instead of mean absolute deviation.

Margaret credited the professional development activities with leading her to some resolution for her long-standing dilemma about why standard deviation is used more often than mean absolute deviation. Teachers worked through an activity in

which they considered the unique minimum produced from a quadratic function in comparison with a function formed from the sum of absolute linear functions.⁷ They also considered the difficulty of working with the sum function versus the quadratic function. At the conclusion of the activity, Margaret proclaimed enlightenment.

I finally know today *why* this works, why we use this one [standard deviation] and not the other one [mean absolute deviation] ... And I've looked it up online and I've seen pages and reams and reams online. That (*points to the activity sheet*) makes more sense than anything.

In her reflections, Margaret proclaimed, “understanding why MAD is not used was a wonderful revelation.” Prior to attending the professional development program, Margaret spent considerable time and energy researching and reading “at least 50 pages” to determine why standard deviation is used and preferred over mean absolute deviation. Prior critical reflection on premises related to measures of variation may have positioned Margaret and others to develop new insights from engaging in professional development activities with others—insights that they may not have formed without prior reflection.

Even if teachers were not successful in resolving their dilemmas by engaging in rational discourse and reflecting on activities during the 40 hours of the professional development program, professional development activities provided a starting point for future reflection. For instance, Caroline indicated that with regard to standard deviation, “I understand it a little bit better but I don't think I can articulate standard deviation.” However, she cites the professional development materials as one resource that she could draw on to further her understandings: “I would...go back and look at my notes.” Although not satisfied with her knowledge about standard deviation or even mean absolute deviation, she suggested that she has confidence that she can enhance her knowledge using the tools she has available to her.

I do like the activities that allow me to at least have some jumping, you know something to jump off of as opposed to looking in the textbook. Cause I liked all of those [professional development] activities. So it [the professional development program] has forced me to be more cognizant of what standard deviation is even though I can't really explain it as well as others.

⁷In statistics, we often try to minimize error from a model for relationships in data. In the case of univariate data, we might consider a model that best represents the data to be a single value, say x , and look to minimize errors, or deviations, from x . Consider a simple data set of two upper arm circumferences of 31 and 44 and two functions to find one or more values of x that minimize error for this data set. A function for the sum of absolute deviations from x could be written as $f(x) = |x - 31| + |x - 44|$. The function for the sum of squared deviations from x could be written as $g(x) = (x - 31)^2 + (x - 44)^2$. The graph of $f(x)$ achieves a minimum for all x between 31 and 44, inclusive, and thus does not yield a unique value for x . The graph of $g(x)$ is a parabola that achieves a minimum at its vertex when x equals 37.5, which also is the mean for 31 and 44. Thus, $g(x)$ yields a unique value at which the sum of squared deviations is minimized, whereas $f(x)$ yields infinitely many values at which the sum of absolute deviations is minimized and not a single well-defined representative value for the data. The graphs become more complicated when considering samples of larger size, but the function for the sum of absolute deviations fails to produce a unique minimum value no matter what sample size is given.

Although Caroline does not explicitly mention reflection as crucial for her future learning, her suggestion that she would need to look back on her notes and professional development activities as a starting point suggests that she recognizes a need for reflection to “become more familiar” and develop deeper understandings of the measures. Other teachers expressed similar sentiments to suggest that they may experience future enlightenment about content from their professional development experiences.

The professional development program activities also provided a starting point for teachers to question and reflect on the strategies they previously used to teach statistics content. In their anonymous feedback on the professional development program, several teachers identified areas they might change in their practices for teaching measures of variation based upon their experiences from the professional development program such as introducing boxplots in conjunction with dotplots when teaching interquartile range or using visual representations to teach mean absolute deviation (to show “visual deviation”). Others mentioned general areas of consideration that likely would impact their teaching about measures of variation. For example, Landen suggested a need to introduce content as more than symbols, examining the content “less algebraically because it is better sometimes to think conceptually.” Bryce identified pushing students to consider premises in addition to content and process as one of the changes he likely would institute in his practice. During his interview, he recounted how some of the teachers struggled to articulate premises as they worked through professional development activities and how their struggles caused him to rethink his practice.

What is it?...I'm doing this, but I don't know what I'm doing. Like, I can get it right...but if you ask me why I did what I just did, I can't tell you...That was really powerful to me...Why do we do what we're doing rather than are you really good at doing it...it's changed my perspective entirely...The way I look at it has changed a lot.

11.6 Discussion

In response to how dilemma, critical reflection, and rational discourse affected teachers' reasoning about and understandings of measures of variation, we first found that professional development activities triggered dilemmas by incorporating standard and nonstandard measures and representations and by focusing on conceptual understanding. Some dilemmas were anticipated and triggered purposefully by triggers embedded in professional development materials whereas other dilemmas had not been anticipated. Notably, dilemmas arose for teachers with different backgrounds in statistics, including teachers with considerable prior experiences with statistics and with sophisticated understandings, teachers who were relatively new to statistics by having only completed a single introductory course focused on procedures, and teachers of both middle and high school students. One of the open questions for transformative learning theory is the types of factors or conditions that trigger transformation (Taylor 1997, 2000). This study contributes knowledge about the types

of experiences that can trigger dilemmas for teachers to potentially transform their content knowledge (and pedagogical content knowledge) for measures of variation. For the teachers in this study, dilemmas served to keep them engaged with professional development program content because they were motivated to resolve their dilemmas.

To facilitate teachers' resolution of dilemmas, the professional development activities provided opportunities for teachers to examine multiple representations and to collaborate with teachers and the professional development facilitators to consider alternative perspectives towards resolution. Specifically, teachers worked in groups consisting of both middle and high school teachers. These teachers were responsible for teaching different statistics content, which resulted in middle school teachers being more familiar with mean absolute deviation and high school teachers being more familiar with standard deviation. Teachers' varied experiences allowed alternative perspectives to be shared for new insights to be gained and for new dilemmas to arise from questions to gain understanding. Several teachers commented on the insights they gained from working with teachers who taught at different grade levels. Teachers also mentioned the importance of visual representations, including area representations, for developing perspectives beyond those possible from symbolic representations. Considerable literature supports the importance of using multiple representations to learn mathematics (e.g. Brenner et al. 1997) and underscores the importance of teachers' knowledge of multiple representations for teaching (e.g. Stohl 2005). This study suggests merit in presenting teachers with alternative perspectives through nonstandard representations of content that focus teachers on premises underlying traditional representations of content.

Discourse and scaffolded activities centered on conceptual understanding served to focus teachers on the premises underlying statistical concepts and procedures to clarify their thinking. With respect to measures of variation, teachers mostly developed new or enhanced meaning schemes for the measures, with several high school teachers developing new meaning schemes for mean absolute deviation and middle and high school teachers enhancing their existing meaning schemes for standard deviation. For some teachers, their meaning schemes about standard deviation may have transformed based on initial views of standard deviation as average distance from the mean and subsequent consideration for why standard deviation typically is used in calculations. Many mathematics education researchers extol the importance of reflection for learning (e.g. Goodell 2000; Roddick et al. 2000), but descriptions of reflective practice largely focus on reflections related to content [e.g. "what I learned this week" (Goodell 2000, p. 50)] or process [e.g. "how I learned it" (Goodell, 2000 p. 50)] without considering reflections on premises underlying content or processes (e.g. "why did I learn from this process?"). Results from this study suggest that reflection on premises, indicative of critical reflection, is an important consideration when designing professional development activities to advance teachers' understandings and development of content knowledge.

Although teachers' learning related to measures of variation mostly was limited to developing new meaning schemes or enhancing existing meaning schemes, some teachers professed enhanced meaning schemes and transformed meaning schemes

related to their teaching of measures of variation and teaching more generally. Daphne, for example, suggested that her meaning scheme for teaching boxplots and interquartile range was enhanced when she proclaimed that her students would benefit from interpreting boxplots when they are superimposed over dotplots. Bryce provides evidence that his meaning schemes for teaching may have transformed as a result of experiencing the content as a learner with other teachers and the foci of the professional development activities with which he engaged. These teachers' successes with deepening their statistical understandings (and presumably their pedagogical content knowledge) suggest merit in designing professional development activities with the constructs of dilemma, critical reflection, and rational discourse in mind.

The study also raises questions about teacher education. Each of the teachers participating in the study completed one or more statistics courses as part of their degree programs. Yet, as evidenced by Jackson's comments, not all of the teachers developed conceptual understandings of foundational concepts such as variation by the conclusion of their teacher preparation programs. Observations such as Jackson's raise questions about the effectiveness of current teacher preparation in statistics and the statistical preparation needed for teachers to develop robust understandings and be effective statistics instructors. Reports such as the *Statistical Education of Teachers* (Franklin et al. 2015) offer recommendations for initial teacher preparation, including recommendations that prospective teachers need powerful learning opportunities to develop conceptual understandings of statistics content and need to engage in the statistical problem-solving process in order to develop appropriate habits of mind for doing statistics. Further research is needed, however, to determine the characteristics of programs, courses, and activities that advance the field towards achieving the vision upon which the recommendations are based.

The professional development program described in this paper aligns with the *Statistical Education of Teachers'* recommendations for practicing teachers, which mirror the principles articulated for the preparation of prospective teachers. This program, however, affected the statistical education of a small number of teachers. Research is needed to investigate how results from teachers' engagement with activities such as those described in this paper might be achieved with larger numbers of teachers to affect change on a larger scale. The increasing availability of freely-available online tools such as the Common Online Data Analysis Platform (CODAP, The Concord Consortium 2017) offer promise for using similar types of technology-based activities with teachers enrolled in Massive Open Online Courses (MOOCs; see the chapter from Pratt, Griffiths, Jennings, and Schmoller in this volume) and from less-developed countries. Questions about whether positive results might be observed from such courses, however, need research-informed answers.

Acknowledgements This paper is based upon work supported by the National Science Foundation under Grant Number 1149403. Any opinions, findings, and conclusions or recommendations expressed are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Biehler, R. (2007). Students' strategies of comparing distributions in an exploratory data analysis context. In *Proceedings of 56th Session of the International Statistical Institute*. Lisbon, Portugal.
- Brenner, M. E., Mayer, R. E., Moseley, B., Brar, T., Duran, R., Reed, B. S., et al. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal*, 34(4), 663–689.
- Clark, J. M., Kraut, G., Mathews, D., & Wimbish, J. (2007). *The "fundamental theorem" of statistics: Classifying student understanding of basic statistical concepts*. Unpublished manuscript.
- Cook, S. A., & Fukawa-Connelly, T. (2016). The incoming statistical knowledge of undergraduate majors in a department of mathematics and statistics. *International Journal of Mathematical Education in Science and Technology*, 47(2), 167–184.
- Cranton, P. (2006). *Understanding and promoting transformative learning: A guide for educators of adults*. San Francisco, CA: Jossey-Bass.
- delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4(1), 55–82.
- Finzer, W. (2002). *Fathom dynamic data software [Computer software]*. Emeryville, CA: Key Curriculum Press.
- Franklin, C. A., Bargagliotti, A. E., Case, C. A., Kader, G. D., Scheaffer, R. L., & Spangler, D. A. (2015). *Statistical education of teachers*. Alexandria, VA: American Statistical Association.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, R., et al. (2007). *Guidelines and assessment for instruction in statistics education (GAISE) report: A pre-K–12 curriculum framework*. 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., delMas, R., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 117–148). Mahwah, NJ: Erlbaum.
- Goodell, J. (2000). Learning to teach mathematics for understanding: The role of reflection. *Mathematics Teacher Education and Development*, 2, 48–60.
- Goos, M., Dole, S., & Makar, K. (2007). Designing professional development to support teachers' learning in complex environments. *Mathematics Teacher Education and Development*, 8, 23–47.
- Jacobbe, T., delMas, R., Haberstroh, J., & Hartlaub, B. (2011). *LOCUS: Levels of conceptual understanding in statistics [Measurement instrument]*.
- Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration [Computer software]*. Emeryville, CA: Key Curriculum Press.
- Lann, A., & Falk, R. (2003). What are the clues for intuitive assessment of variability? In C. Lee (Ed.), *Reasoning about variability: A collection of research studies. Proceedings of the third international research forum on statistical reasoning, thinking, and literacy*. Mount Pleasant: Central Michigan University.
- 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., & 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.
- Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal*, 4(1), 27–54.
- Mathews, D., & Clark, J. (2003). *Successful students' conceptions of mean, standard deviation, and the central limit theorem*. Unpublished manuscript.
- Mezirow, J. (1991). *Transformative dimensions of adult learning*. San Francisco: Jossey-Bass.
- Mezirow, J. (2000). Learning to think like an adult: Core concepts of transformation theory. In J. Mezirow & Associates (Eds.), *Learning as transformation: Critical perspectives on a theory in progress* (pp. 3–34). San Francisco, CA: Jossey-Bass.

- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards in Mathematics*. Washington, DC: Council of Chief State School Officers.
- Peters, S. A. (2011). Robust understanding of statistical variation. *Statistics Education Research Journal*, 10(1), 52–88.
- Peters, S. A. (2014). Developing understanding of statistical variation: Secondary statistics teachers' perceptions and recollections of learning factors. *Journal of Mathematics Teacher Education*, 17(6), 539–582.
- Peters, S. A., & Kopeikin, K. S. (2016). Integrated reasoning about statistical variation: Secondary teachers' development of foundational understandings. In D. Ben-Zvi & K. Makar (Eds.), *The teaching and learning of statistics* (pp. 249–259). Cham, Switzerland: Springer.
- 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.
- Roddick, C., Becker, J. R., & Pence, B. J. (2000). Capstone courses in problem solving for prospective secondary teachers: Effects on beliefs and teaching practices. In T. Nakahara & M. Koyama (Eds.), *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 97–104). Hiroshima: Hiroshima University.
- Sánchez, E., Silva, C. B., & 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: A joint ICMI/IASE study* (pp. 211–221). Dordrecht, The Netherlands: Springer.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester Jr. (Ed.), *Handbook of research on mathematics teaching and learning* (2nd ed., pp. 957–1009). Greenwich, CT: Information Age.
- Silva, C. B., & Coutinho, C. Q. S. (2008). Reasoning about variation of a univariate distribution: A study with secondary math 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, Mexico: International Commission on Mathematical Instruction and the International Association for Statistical Education. Retrieved October 12, 2008, from http://www.ugr.es/~icmi/iase_study/.
- Smith, T. M., Desimone, L. M., & Ueno, K. (2005). “Highly qualified” to do what? The relationship between teacher quality mandates and the use of reform-oriented instruction in middle school mathematics. *Educational Evaluation and Policy Analysis*, 27(1), 75–109.
- Sorto, M. A. (2004). *Prospective middle school teachers' knowledge about data analysis and its applications to teaching*. Unpublished doctoral dissertation, Michigan State University, Lansing.
- Stohl, H. (2005). Probability in teacher education and development. In G. A. Jones (Ed.), *Exploring probability in schools: Challenges for teaching and learning* (pp. 345–366). New York: Springer.
- Taylor, E. W. (1997). Building upon the theoretical debate: A critical review of the empirical studies of Mezirow's transformative learning theory. *Adult Education Quarterly*, 48(1), 34–60.
- Taylor, E. W. (2000). Analyzing research on transformative learning theory. In J. Mezirow and Associates (Eds.), *Learning as transformation: Critical perspectives on a theory in progress* (pp. 285–328). San Francisco, CA: Jossey-Bass.
- The Concord Consortium. (2017). Common online data analysis platform (Version 2.0) [Mobile application software]. Retrieved from <https://concord.org/>.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Erlbaum.
- 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. (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. (2005). The winds are variable: Student intuitions about variation. *School Science and Mathematics*, 105(5), 252–269.
- 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. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society*, 174(2), 247–295.

Chapter 12

Exploring Secondary Teacher Statistical Learning: Professional Learning in a Blended Format Statistics and Modeling Course



Sandra R. Madden

Abstract Providing opportunities for secondary teachers to develop the statistical, technological, and pedagogical facility necessary to successfully engage their students in statistical inquiry is nontrivial. Many mathematics and science teachers in the U.S. have not benefitted from sufficient opportunity to learn statistics in a sense-making manner. With statistics assuming a more prominent place in the secondary curriculum, it remains a priority to consider viable ways in which to reach and support the statistical learning trajectory of both pre- and in-service teachers. This study explores ways in which a course that blends face-to-face and virtual learning experiences impacted in-service teachers' technological pedagogical statistical knowledge (TPSK). Results suggest the course positively impacted participants' TPSK.

Keywords Blended learning · Professional development · Statistics · Teaching

12.1 Introduction

Statistics has achieved a position of status in the Pre-K-12 curriculum in the United States and around the world; (Australian Curriculum Assessment and Reporting Authority 2010; National Governors Association Center for Best Practices & Council of Chief State School Officers 2010; Conference Board of the Mathematical Sciences 2010; Franklin et al. 2007). Secondary mathematics teachers are now responsible for teaching statistics; yet remain ill prepared for the job (Batanero et al. 2011; Conference Board of the Mathematical Sciences 2010; Franklin et al. 2015; Madden 2008; Shaughnessy 2007). In contrast to the largely theoretical statistical courses teachers tend to take in mathematics departments, recent recommendations suggest the need for authentic data-intensive exploration and modeling experiences in addition to theory-based coursework (Franklin et al. 2015). Teachers should develop facility with the statistical process (Wild and Pfannkuch 1999); techniques and tools for simula-

S. R. Madden (✉)
University of Massachusetts Amherst, Amherst, MA, USA
e-mail: smadden@umass.edu

© Springer Nature Switzerland AG 2019

G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_12

tion, computation, and representation; and a generally elevated understanding of the statistical landscape appropriate to meet 21st century curricular demands. Courses to prepare teachers for these new demands are still rare and largely unexamined (Franklin et al. 2015).

Related to pedagogical content knowledge (PCK) (Shulman 1986) and technological pedagogical content knowledge (TPCK) (Mishra and Koehler 2006), technological pedagogical statistical knowledge (TPSK) (Lee and Hollebrands 2011) addresses the importance of teachers understanding students' learning and thinking about statistical ideas; conceptions of how technology tools and representations support statistical thinking; instructional strategies for developing statistics lessons with technology; critical stance towards evaluation; and use of curricula materials for teaching statistical ideas with technology (Groth 2007). TPSK informs a *doer to designer* approach (Kadijevich and Madden 2015) to teacher learning where teachers first engage in statistical investigations as learners (doers) and later design, implement and study the implementation of statistical lessons as enacted (designers). Constructionism (Papert 1991) is echoed in the *doer to designer* framework with its emphasis on engaging teachers as statistical learners en route to supporting them to design, implement, and reflect on statistical learning opportunities with their own students.

With these perspectives as guides, a blended format course (part face-to-face, part virtual) was developed to support and explore teachers' evolving TPSK. This study begins to address the dearth of research exploring teachers' TPSK development in relation to the enacted curriculum in the classroom (Kadijevich and Madden 2015; Lee and Nickell 2014).

12.2 Description of the Blended Learning Environment

A three-credit experimental graduate course offered in a US university was designed by the author to facilitate middle and high school teacher learning of statistics and modeling in the secondary curriculum. The course intended to impact teachers' practices and their students' opportunity to engage with statistical ideas. Design commitments included: active learning, technology rich investigations, community of practice orientation (Wenger 1998), exploration of curriculum materials, and attention to autonomy (Ryan and Deci 2000). The course consisted of five face-to-face (F2F) four-hour sessions plus five virtual modules between F2F meetings (Fig. 12.1). This course structure facilitated teachers' schedules with intense statistical and technological learning experiences during their summer break and thoughtful implementation of statistical units of instruction with secondary students when the teachers returned to school in September.

Course content included model-based sampling investigations, experimental design investigations to motivate randomization testing, and other simulation-based statistical tools for supporting statistical argumentation. Face-to-face sessions were largely focused on statistical investigations intended to support the statistical process and conceptual development, use of technology for exploring data, small and

| F2F <i>June</i> | | Virtual Modules <i>July</i> | | | F2F <i>August</i> | | Virtual Modules <i>September</i> | | F2F <i>October</i> |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 1 | 2 | 3 | 3 | 4 | 4 | 5 | 5 |
| Engaging and extending prior statistical knowledge; building community; informal inferential reasoning; intro to <i>TinkerPlots</i> , <i>Fathom</i> , and <i>CPMP-Tools</i> | | Building facility with technology, generating and interrogating data created with the <i>TinkerPlots Sampler</i> , exploring and experiencing curriculum units, analyzing curricula | | | By chance or by cause—experimental design and randomization testing | | Simulating randomization testing with <i>TinkerPlots</i> and <i>Fathom</i> ; learning new statistics through curriculum exploration; planning for unit implementation | | Sharing investigative & implementation experiences; empirical sampling distributions, Central Limit Theorem (CLT), regression and correlation |

Fig. 12.1 Course format, schedule, and content trajectory

whole group processing of readings and experiences, and general community building (Fig. 12.2).

In addition, participants were invited to read approximately 30 articles, conduct a statistical curriculum analysis, and engage in an action research project in which they designed, implemented, and reflectively analyzed student learning in a statistical unit of study where technology was utilized. Participants electronically submitted written assignments and discussion posts for each of the 10 distinct chunks of the course. Appendix A (<http://bit.ly/2OAKugq>) provides an example of instructions for participants for one of the virtual modules. Appendix B (<http://bit.ly/2OAKugq>) provides a description of the curriculum analysis project and associated scoring rubric. Appendix C (<http://bit.ly/2OAKugq>) contains the instructions for the curricu-

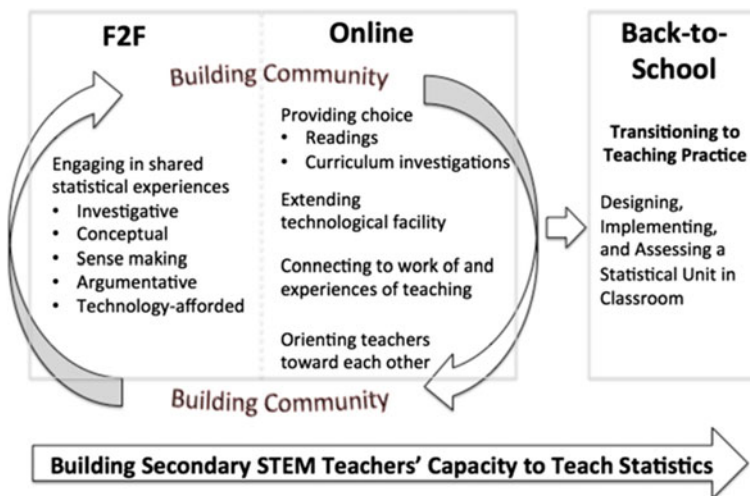


Fig. 12.2 Process-related design commitments

lum implementation project. Course grades were determined by 40% preparation and participation; 30% curriculum analysis project, 30% action research project (curriculum implementation).

Aspects of the design of the course for participants included: (1) developing facility with *Fathom* (Finzer 2005), *TinkerPlots* (Konold and Miller 2005), and *CPMP-Tools* (Keller 2006) software while conducting statistical investigations, much of this during virtual modules; (2) analyzing secondary-level curriculum materials to support statistical development as well as pedagogical sensibility; (3) designing, implementing, and studying a technologically-relevant statistical unit in their own classroom; and (4) choosing articles to read and statistical content and curricula to investigate from a pool of recommendations. The author provided a library of curriculum materials and literature for this study.

12.3 Methods

Ten secondary teachers (eight mathematics, two science) participated in the study. Four of these participants were Teaching Fellows in a National Science Foundation (NSF)-funded Noyce Master Teaching Fellow/Teaching Fellow project, while six were volunteers from schools not associated with the Noyce project. All names are pseudonyms. Each participant completed an initial background and motivation survey as well as post course survey (see <https://goo.gl/forms/gaBXAPkFOzTBW1XP2>). All course assignments, discussions, emails, and associated artifacts were collected for analysis. Survey data were analyzed using descriptive statistics and standard quantitative methods. Document analysis techniques were used for qualitative data with open and axial coding. With a focus on teachers' development of TPSK, initial codes included: statistical knowledge (SK), technological knowledge (TK), pedagogical knowledge (PK), STK, SPK, TPK, TPSK, tool use, impact of curriculum, impact of activity, impact of reading, impact of discussion, challenges, and miscellaneous. Each data source (e.g., discussion post or written assignment) was analyzed and summarized. Coding categories were further explored for themes across the data. Data were analyzed vertically by type and horizontally by person. A chronological case study analysis for each participant was conducted to capture the evolution of each participant's learning over the period of the course to answer the research question: *To what extent and in what ways did the blended format statistics and modeling course experiences impact participants' TPSK?*

A portion of the analysis is reported in this study. Results will coordinate teachers' self-reported data with data analyzed by the researcher. Changes in teachers' perceptions of statistical and technological facility are summarized; descriptions showcasing the breadth of curricular investigations and implementation projects are presented; and two specific learning trajectories are provided to illustrate the development of TPSK for project participants.

12.4 Results

12.4.1 Analysis of Participants' Self-reported Pre- and Post-intervention Data

An analysis of participants' comfort level (1-low, 5-high) with statistical big ideas pre- and post-intervention suggests limited prior statistical knowledge for most and significant improvement in a number of areas (Table 12.1). Significant gains in the areas of descriptive statistics, experimental design, sampling distributions, overall, and facility with *TinkerPlots* and *Fathom* coincide with the goals of the course (see Table 12.2). Understanding of statistical graphs showed improvement but was also the area most highly rated during the initial survey, and gains scores were not significant. Correlation and regression were addressed only briefly at the end of the course; however, some participants elected to explore curriculum units where these were a focus. This decision to focus elsewhere was predicated on the fact that many secondary teachers tend to have some familiarity with regression and correlation through their work teaching algebra. Several participants selected instructional units addressing correlation and regression during one of the modules where they could choose from a variety of statistical units to explore. The relatively high standard deviation associated with correlation and regression may be the result of representing a bifurcation of experiences where some participants benefitted from independent work, while others did not. Statistical inference was the area seeing the least change, a result likely due to the more informal approach to inference that participants may not have associated with more formal statistical inference.

Participants rated their personal engagement in the course (e.g., course readings, statistical tasks and investigations, discussion posts, curriculum units, *TinkerPlots*, *Fathom*, *CPMP-Tools*). Aggregate ratings (1-low, 4-high) ranged from 2.86 to 3.71 ($M=3.33$, $SD=0.31$) and were strongly, positively associated with perceived learning gains (Fig. 12.3).

Participants rated the extent to which course objectives were met. Mean ratings (1-low to 5-high) were 4.30 or above with five of seven objectives receiving a median rating of 5 (Table 12.2), suggesting participants believed course objectives were met.

12.4.2 Curriculum as Lever to Promote TPSK

Curriculum played a major role in the course. Curriculum frameworks such as GAISE and CCSSM were introduced to participants. Innovative curriculum texts developed with funding from the NSF such as *Core-Plus Mathematics Project (CPMP)* (Hirsch et al. 2015), Interactive Mathematics Program (IMP) (Fendel et al. 2012) and *Connected Mathematics Project (CMP)* (Lappan et al. 2009) were utilized to develop statistical ideas as well as to introduce participants to innovative instructional materials. These materials allowed for modeling classroom instruction in a manner that

Table 12.1 Participants' self-reported comfort level with statistical big ideas and tools (1-low, 5-high)

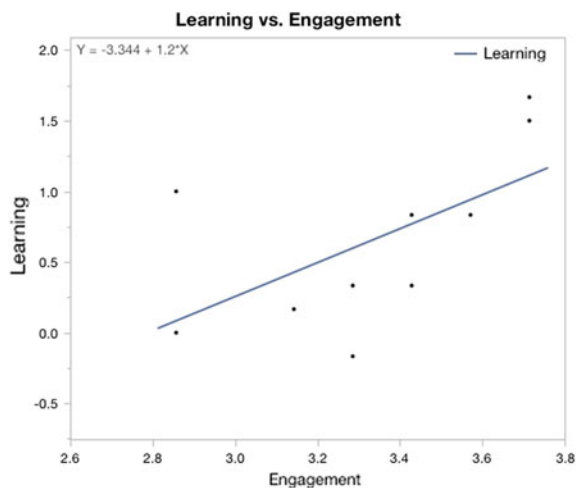
| | Mean | | Median | | SD | | Gain Score (Means) | | p-value |
|----------------------------------|------|------|--------|------|------|------|--------------------|------|---------|
| | Pre | Post | Pre | Post | Pre | Post | Mean | SD | |
| N=10 | | | | | | | | | |
| Descriptive statistics | 3.00 | 3.90 | 3 | 4 | 0.47 | 0.74 | 0.90 | 0.57 | 0.0039* |
| Statistical graphs | 3.90 | 4.40 | 4 | 5 | 0.88 | 0.84 | 0.50 | 0.85 | 0.0917 |
| Experimental design | 2.90 | 3.60 | 3 | 4 | 1.10 | 0.84 | 0.70 | 0.82 | 0.0289* |
| Correlation and regression | 2.30 | 2.80 | 2.5 | 3 | 1.34 | 1.55 | 0.50 | 1.18 | 0.1919 |
| Sampling distributions | 1.80 | 2.80 | 2 | 3 | 0.79 | 0.79 | 1.00 | 1.05 | 0.0140* |
| Statistical inference | 2.00 | 2.30 | 2 | 2 | 1.05 | 0.95 | 0.30 | 0.95 | 0.2622 |
| Overall (aggregate ratings) | 2.65 | 3.30 | 2.5 | 3.5 | 0.74 | 0.71 | 0.65 | 0.62 | 0.0019* |
| Facility with <i>TinkerPlots</i> | 2.10 | 3.70 | 2 | 3.5 | 1.29 | 0.82 | 1.60 | 1.26 | 0.0054* |
| Facility with <i>Fathom</i> | 1.40 | 3.30 | 1 | 4 | 0.70 | 1.06 | 1.90 | 0.99 | 0.0017* |

*Represents gain scores statistically significantly greater than 0, with $p < 0.05$

Table 12.2 Ratings for the extent to which course objectives were met (1-low, 5-high), $N = 10$

| Course objectives | Summary ratings |
|--|----------------------------------|
| To explore issues of secondary mathematics curriculum recommendations and standards, curriculum design, curriculum implementation, and curriculum research | $M = 4.60, Mdn = 5, SD = 0.70$ |
| To support understanding of important curricular trends and innovations in statistics education at the middle and high school level | $M = 4.30, Mdn = 4, SD = 0.67$ |
| To support understanding of and relationships among important statistical big ideas, most notably, distribution, variability, and sampling distributions as they relate to comparing distributions | $M = 4.60, Mdn = 5, SD = 0.70$ |
| To support statistical reasoning, thinking, and literacy, generally | $M = 4.70, Mdn = 5, SD = 0.48$ |
| To develop facility with innovative technological tools for exploring data and conducting statistical analyses. | $M = 4.70, Mdn = 5, SD = 0.48$ |
| To become familiar with research in the area of statistics education in order to critically examine curricular implications with respect to statistical reasoning, thinking, and literacy | $M = 4.70, Mdn = 5, SD = 0.48$ |
| To support ability to implement high quality statistical instruction at the secondary level | $M = 4.40, Mdn = 4.5, SD = 0.70$ |

Fig. 12.3 Participants' self-reported statistical learning (scale 1-5) versus self-reported course engagement (scale 1-4)



privileged investigation, discovery, and argumentation. Learning that instructional materials like those used during the course existed helped to encourage participants to critically examine them. The curriculum analysis project allowed participants to look

Table 12.3 Descriptions of participants' curriculum analysis projects

| Teacher | Grade Level | Project |
|--------------------|-------------|---|
| <i>Mathematics</i> | | |
| Ingrid | Middle | Analysis of Connected Mathematics Project (CMP) 2 and Big Ideas Math Sixth Grade Curriculum using GAISE Frameworks |
| Karen | Middle | Comparing Connected Mathematics Project 2, Data About Us to Big Ideas Math, Grade 6, Chaps. 9 and 10 |
| Alexia | Middle | Comparing Connected Mathematics Project, Samples and Populations and 7th and 8th grade Big Ideas Math |
| Shelley | High | Exploring Core-Plus Mathematics Project (CPMP) Units with GAISE Framework for use in College Prep Statistics Course |
| Jared | High | Comparing the Interactive Mathematics Program (IMP), Game of Pig, with The Basic Practice of Statistics (BPS), Chap. 4 Probability and Sampling Distributions |
| Claire | High | Comparing Interactive Mathematics Project (IMP), Pit and the Pendulum Unit to Carnegie Learning Algebra 1, Chap. 8 |
| Joanna | High | Comparing Project Lead the Way (Statistics 4.1) to Core-Plus Mathematics Project, Unit 2, Lesson 1 |
| Trevor | High | Curriculum Analysis of Core-Plus Mathematics: Contemporary Mathematics in Context Courses 1 & 2 |
| <i>Science</i> | | |
| Alexandra | High | Exploring the development of standard deviation using Core-Plus Mathematics Course 1 Unit 1 and χ^2 using Transition to College Mathematics Unit 1 |
| Michelle | High | Coordinating GAISE Framework, A.P. Quantitative Skills-A-Guide for Teachers, The Handbook for Biological Statistics, Using BioInteractive Resources to Teach Mathematics and Statistics in Biology, AP Biology: Course Description, Next Generation Science Standards, and Understanding by Design (UbD) Unit Review rubric from Department of Elementary and Secondary Education |

carefully at ways in which different curriculum materials have potential to engage students in statistical activity as well as to address state and national standards. Contrasts with more familiar materials became obvious. As Table 12.3 illustrates, all mathematics participants and one science participant elected to analyze some combination of curriculum materials that included NSF-funded materials. The other science participant selected a broad range of resources for Advanced Placement (AP) Biology (College Board 2015) to examine and critique. Completed projects were posted on Moodle for sharing and brief presentations were made during a F2F session.

By the end of the course and as will be illustrated in Sects. 12.4.3 and 12.4.4, participants developed and demonstrated extensive familiarity with the *GAISE* Framework and several high quality instructional resources for supporting student statistical learning. They increased their facility with the use of dynamic statistical tools (e.g., *TinkerPlots*, *Fathom*, and *CPMP-Tools*) as they engaged in statistical activity as learners. The requirement to complete curriculum implementation action research projects at the end of the course signaled the expectation that lessons learned would be explicitly tied to classroom practice. Using action research methods, participants designed, implemented, and analyzed student learning from a statistical unit of study. Table 12.4 contains brief descriptions of participants' focus for their project and the technological tool(s) they elected to implement with students.

A wide range of statistical content was addressed and explored through the curriculum implementation projects; however, it was essential that participants could select appropriate content for their particular teaching context. Participants briefly presented their projects on the final day of the course. Their unique and improved statistical, technological, and pedagogical knowledge was evidenced through these individual projects and will be further described throughout the next two sections.

12.4.3 Tracing Learning Trajectories: Examining Two Cases for TPSK

Tracing participants' learning journeys over the course illuminated a complicated but compelling storyline for each participant. Every participant attempted and completed all aspects of the course; however, the extent to which each aspect was completed varied considerably. Only a tiny fraction may be presented here, so I illustrate trajectories of two distinct patterns of engagement.

12.4.3.1 The Case of Claire

Claire is a third year high school mathematics teacher who described her past largely theoretical statistical learning experiences in great detail and characterized them as procedurally dominated:

I took a 1.5 credit Prob Stats course on: Sample spaces, events, axioms for probabilities; conditional probabilities and Bayes' theorem; random variables and their distributions, discrete and continuous; expected values, means and variances; covariance and correlation ... Also, I've taken a 2 credit Intermediate Probability course on: Continuous random variables, distribution functions, joint density functions, ... Chebyshev' theorem ... Most of the class time was spent taking notes in a "fill in the blank" format and then once in a while we had statistical investigations. The professor did not take time to know her students individually and I felt that I didn't learn much in her class because of this.

She indicated a desire to "learn methods for teaching statistics in a meaningful and engaging manner." Her pre-course statistics comfort level was 2.33.

Table 12.4 Participants' statistical curriculum implementation projects and associated technological tool

| Teacher | Grade level | Curriculum implementation project | Technology utilized with students |
|-----------------|-------------|--|-----------------------------------|
| Ingrid | Middle | What strategies and/or tools do students employ in reasoning about best measure of center to use to describe a data distribution? How is this reasoning explained in light of the intended and implemented curriculum? <i>CMP2, Data Distributions</i> | <i>TinkerPlots</i> |
| Karen | Middle | To what extent does students' use of multiple representations through dynamic software and exploratory work impact student understanding on the effect outliers have on the mean? | <i>TinkerPlots</i> |
| Alexia | Middle | How does qualitative graphing of data help students understand the concept of slope? | <i>TinkerPlots</i> |
| Jared & Shelley | High | How does performing a randomization distribution on student data impact student's understanding of statistical likelihood? | <i>Fathom</i> |
| Claire & Trevor | High | How does having a statistical context for a problem support and/or impact student understanding of lines of best fit? | <i>TinkerPlots</i> |
| Joanna | High | Statistics for 9th grade mathematics enrichment | <i>Excel, CPMP-Tools</i> |
| Alexandra | High | Introducing standard deviation with CPMP Course One, Unit 2, Lesson 2, Investigation 4, "Measuring Variability: The Standard Deviation." Students will demonstrate their learning through a summative assessment and by analysis of lab data from the "Rainbow Osmosis" lab. | <i>CPMP-Tools</i> |
| Michelle | High | Statistics in AP Bio: Scaffolding Student Understanding for HHMI Biointeractive Curriculum | <i>Fathom</i> |

Following the June F2F sessions and readings, her reflection, a portion of which is below, indicated her growing understanding of the use of graphing calculators, *TinkerPlots*, and the simulation process model for generating empirical sampling distributions:

In Using Graphing Calculator Simulations in Teaching Statistics, Koehler gives a pretty detailed description of how to use the graphing calculators, and I realize that the graphing calculators are much more powerful than even I knew. However, I found that this tool is much

more syntactically confusing and I would anticipate that students would have a lot of trouble understanding what was truly happening in situations being modeled. In contrast, Lane-Getaz describes that *Tinkerplots* really allows students to see the three layers of statistical modeling with a great figure on page 280 of the yearbook. I think I finally have this whole process clear in my mind! Finally, Lane cautions teachers that simulations can sometimes still produce passive learners, so they must be presented with a query-first method of teaching. I really want to remember this idea and try to pose a question of study to my students at the beginning of units and lessons of study.

In July, she assessed her own understanding after reading the *Guidelines for Assessment and Instruction in Statistics Education* (Franklin et al. 2007) using 1, 2, and 3 for levels A, B, and C:

I think I am probably around level 2.5, if we're allowing halves. I've heard of some level 3 concepts, but do not have a firm grasp on, for example, the data analysis done on pp 67–70. The coolest new thing that I learned about was the Quadrant Count Ratio. I didn't know there were more than one "correlation coefficient" although in retrospect it makes sense that there isn't just one. I like that I now could explain how to find this one, whereas I still have no idea how Pearson's correlation coefficient is calculated.

During her curriculum analysis project she compared a unit from the *Interactive Mathematics Project (IMP)* to a unit developing similar content (standard deviation) from her school's newly adopted *Carnegie Learning Program*. She concluded *IMP* provided more cognitively demanding tasks for students, but both texts performed equally when compared to *GAISE* recommendations.

In September following a series of readings and tasks supporting understanding the randomization test for comparing experimental treatment and control groups, she writes about her own growth with *TinkerPlots* and *Fathom* and compares to *CPMP-Tools*:

I think that TP and Fathom allow for a deeper understanding than *CPMP tools* because you are building more of the functionality yourself. You have to work directly with the resampling process, so you understand exactly what is happening and how the means are being calculated. I understand better now how to use formulas in Fathom, and am gaining ability with Fathom. I haven't used it much before, but this is the second assignment I've completed with it. I'm improving at using the sampler in TP.

In October, she attributes improved understanding of binomial distributions to her reading selection.

'Is Central Park Warming?' This article describes an activity that students can do to find out the probability that the warm temperatures in Central Park happened randomly. They then compare this to the exact mathematical probability calculated from the binomial distribution. This provided some insight to me about what the binomial distribution actually is!

For her curriculum implementation project, Claire partnered with a classmate to design and implement a statistical unit in her peer's class. Together, they developed and reflected on the unit, its implementation, and impact on student learning. She was unable to implement a statistics unit with her own classes due to curricular limitation within the window for the course, so this partner project allowed her to still design and study the implementation for the purpose of the course.

In the final survey, she remarked,

This course has exposed me to literature in the field of statistics education which I can bring to other educators in my school. I understand the flow of statistical learning that should happen in middle and high schools. I think the most important pedagogical idea that I have taken away is that it is more important for students to construct and use their own measures in statistics before learning about and applying conventional measures. I very much feel like I have more resources for the future.

Her comfort level with statistical ideas jumped to 3.83 at the end of the course and her perceived facility with *TinkerPlots* and *Fathom* increased from 3 to 5 and 1 to 4, respectively. Claire is a case of a teacher from a highly regarded undergraduate institution with a bachelor's degree in mathematics, master's degree in education, and prior to her taking the course described herein, a very fragile understanding of statistics with few constructive ways in which to teach statistics. Throughout the course, she engaged thoroughly in tasks, investigations and all assignments and her written record indicates strong growth as a learner and teacher of statistics; that is, her TPSK improved dramatically. She communicates growing sensibilities about statistics as a discipline, teaching statistics in a learner centered, technologically oriented manner and alignment with professional guidelines for teaching.

12.4.3.2 The Case of Alexandra

Alexandra is a veteran high school science teacher who wrote, "I had a statistics course in college...many years ago. I have been teaching chi square and standard deviation to AP Biology students as part of the newest version of the course and feel I need more background." Her overall pre-course statistical comfort level was 1.67. Following the initial F2F sessions and Module 1, Alexandra wrote,

Learning takes time, and good instruction loaded with experiences for students to develop their own understanding takes LOTS of time ... I was impressed (overwhelmed?) by the topics listed in the Common Core for the Statistics & Probability strand. To me, even the Grades 6/7/8 expectations seemed very challenging. I thought the detailed descriptions and examples for Levels A/B/C as detailed in the GAISE Report were very helpful. I especially liked how in some cases the same activity or exercise was used at multiple levels, to distinguish the differences in understanding expected.

Following Module 2, she continued to express a sense of excitement, challenge and pedagogical insight related to her activity:

After using *TinkerPlots* myself, I don't need the experts to convince me of how helpful this software tool could be in my classroom. However, extensive time in a computer lab is difficult to schedule in my school, and finding extensive time for any new activity is a challenge! I will explore using *TinkerPlots* to some degree, but what I found most interesting and potentially useful in this set of readings was the exercise described by delMas and Liu in "Exploring students' conceptions of the standard deviation." I can see how I could use the pairs of graphs on page 62 (they call them test items) to help my students understand standard deviation. In the study, students were asked to decide if, for each pair, the second graph would have a higher or lower standard deviation than the first. By predicting, calculating/confirming, and discussing these pairs of graphs 1 or 2 at a time, I believe my students could develop a better understanding of standard deviation.

For her curriculum analysis, Alexandra chose to explore two units from *Core-Plus Mathematics Program (CPMP)* (Hirsch et al. 2015), one focusing on standard deviation and the other on the χ^2 test. Due to the mathematical demands of the χ^2 unit, she sought out and discovered additional AP Biology (College Board 2015) resources to support her learning that she shared with the other science teacher in the course. The curriculum analysis project allowed her to build her own capacity to understand and teach two important statistical ideas to her students.

Following Module 4, she demonstrated her grasp of randomization testing and *TinkerPlots* facility:

I really had to follow the videos closely to do the randomizations initially, and even then I needed additional assistance (Thanks person1 and person2!). But I just corrected a quiz for my AP class...2 versions, means of 13.0 and 13.5. I was able to run a randomization test using *TinkerPlots* to confirm that the difference in the quiz means has p value of 0.63, so I think I can tell the students that one quiz was not easier than the other! While this (Module 4) was time-consuming, between the exercises from the unit, the videos, the software practice, and the readings, I feel very confident about my understanding of and my potential use of/teaching of these concepts/tests.

Alexandra's curriculum implementation was exemplary. She presented thoughtful plans to build ideas of standard deviation with her students, used the *CPMP* unit from her curriculum analysis project and utilized *CPMP-Tools* with her students. She videotaped her classroom, collected student artifacts, and reflected on the experience with a colleague. Her reported insights showed her vulnerability as well as her strengths as a teacher and champion for students. Alexandra's project illuminated her growth in statistical knowledge, technological statistical knowledge, her student's growth in statistical knowledge, and ultimately markedly improved TPSK. She indicated a disposition toward continuing to grow and learn in this arena.

On the post-course survey, she wrote:

I learned a lot about statistical concepts and tools. I learned a lot about how students learn statistics. This will have a direct impact on my classroom and my students, as I am better prepared to help them understand measures of central tendency, variation, standard deviation, p values, and chi square. I benefited from the exposure to technological tools, but could use a lot more practice to feel truly comfortable using them. I learned about issues, challenges, and successes that other teachers have in teaching statistical content to students. I learned a lot from being a student and working in groups with others in completing some of the exercises. I feel even more strongly that students need to understand the concepts behind the statistical tools (what do they mean?). I have a much better sense of how the tools can be applied to our own data sets. I found the exercises that we completed in class in groups to be excellent learning activities in terms of concepts but also as models of teaching strategies. I enjoyed working through the *CPMP* Lessons; I really like their approach in introducing concepts gradually and before the equations and/or technological aids. They include pertinent examples and plenty of practice problems. I enjoyed working with *TinkerPlots* and *Fathom* and am convinced of their power in illustrating many statistical concepts (randomization, value of large data sets). Great experience. I would not have signed on if it had been offered online only. The face-to-face sessions were particularly beneficial, and I believe the online modules worked better given that we knew who the other students were when posting comments, questions, etc.

Her post-course statistical comfort level was 2.5. This rating seems to confirm her awareness of the complexity of statistical learning, but perhaps underestimates her actual learning. It may well illustrate this teacher's acknowledgement of learning while also recognizing the need to learn more. She seemed to recognize a state of personal disequilibrium while at the same time developing agency in the statistical teaching and learning realm. Alexandra's reaction to the blended format of the class suggests a real preference for face-to-face interaction to build community and it foregrounds potential reasons why hybrid statistics courses with face-to-face and virtual components may support better learning for students than online only experiences (Meyer and Lovett 2014).

12.4.4 Summary Perspectives Across Participants

Each of the other eight participants' individual storylines vary, yet they each demonstrated improved TPSK. As Table 12.5 illustrates, eight of 10 participants assessed their statistical knowledge to have increased. One student's rating from pre- to post did not change; however, from the perspective of the instructor, this student demonstrated increased statistical knowledge. During his curriculum implementation project, he designed a set of lessons to introduce his students to randomization testing for comparing results from two groups in an experimental context. Randomization testing as a means for comparing experimental and treatment groups was unfamiliar to all participants prior to the course, thus this represents significant growth in statistical knowledge. The participant with a negative gain score represents a student who demonstrated remarkable engagement with all aspects of the course and a growing facility with statistical ideas and tools; however, the student may not have felt completely competent yet. As the participant mentioned in her curriculum implementation project,

It [the course] benefitted me by giving me an awareness of statistical learning and concepts at the high school level. Some of the concepts we learned about I don't think I realized were of the statistical realm. It just made me realize that there is so much I don't understand and I feel like a novice. The course just really gave me an awareness that statistics is different than math and I need to approach it differently with my students (Michelle).

The TSK scores in Table 12.5 represent self-reported gain scores with *Fathom* and *TinkerPlots*. As the data show, every participant increased their TSK with at least one technology. The two participants whose gains were zero or negative had rated their facility highly on the initial survey and likely discovered there was much more to learn than they had realized. As indicated in Table 12.1, gain scores for both *Fathom* and *TinkerPlots* were significantly greater than 0.

Finally, each participant's completed curriculum implementation project provided evidence of TPSK growth. Three levels of TPSK were evident through the projects. At the lowest level (✓), projects fell into one of three categories: (1) largely algebraic reasoning rather than statistical reasoning but utilized technology productively; (2)

Table 12.5 Summary of participants’ self-reported statistical understanding pre- and post-intervention gain scores as the self-reported average gain in facility with *TinkerPlots* and *Fathom*, and instructor assessment of TPSK demonstrated through curriculum implementation projects

| Teacher | Pre | Post | SK pre → post | TSK gains pre → post Fathom/TP | | | TPSK Curriculum implementation project | Course grade |
|-----------|------|------|------------------|--------------------------------------|----|---|---|-----------------|
| | | | | F | TP | | | |
| Alexandra | 1.67 | 2.50 | + | 1 | 2 | + | ++ | A |
| Ingrid | 2.00 | 3.67 | + | 2 | 2 | + | ++ | A |
| Jared | 2.17 | 2.17 | 0 | 2 | 1 | + | + | A– |
| Claire | 2.33 | 3.83 | + | 3 | 2 | + | + | A |
| Alexia | 2.33 | 3.33 | + | 3 | 1 | + | ✓ | A |
| Michelle | 2.67 | 2.50 | – | 3 | 2 | + | ++ | A |
| Joanna | 2.83 | 3.17 | + | 0 | 1 | + | ✓ | A– |
| Trevor | 2.83 | 3.67 | + | 2 | 4 | + | + | A– |
| Karen | 3.5 | 3.83 | + | 1 | –1 | + | + | A |
| Shelley | 4.17 | 4.33 | + | 2 | 2 | + | ✓ | B+ |

✓ indicates evidence of beginning TPSK, + indicates evidence of strong TPSK, ++ indicates evidence of excellent TPSK

relied on previously familiar technology and content but incorporated more student-centered activity; and (3) relied heavily on a partner to do the technological or statistical heavy lifting. At the (+) and (++) levels, participants’ projects showcased greater evidence of stretching in the direction of engaging learners with less familiar content using tools and materials that were initially unfamiliar. Projects rated (++) were exceptional and represented thoughtful and thoroughly documented and analyzed products. Two of the three projects in this category were from science teachers.

Document analyses further supported the following claims: (1) science teachers in this environment appeared unusually receptive to learning statistics and adapting their learning to their practice; (2) teachers with the highest self-reported statistical comfort level tend to be those with significant statistics teaching experience and least receptive to new ideas; (3) modeling using resampling ideas such as randomization testing in technologically-conducive environments is accessible and beneficial; (4) analyzing curriculum materials using *GAISE* (Franklin et al. 2007), National Council of Teachers of Mathematics (NCTM 2000, 2009), Next Generation Science Standards (National Research Council 2013), and Advanced Placement Biology (College Board 2015) guidelines is worthwhile for teachers; and (5) pushing for teachers to design, implement, and reflect on students’ statistical learning is formidable yet impactful.

12.5 Discussion

Creating experiences with potential to directly impact participants' capacity to design, implement, and reflect on statistical units in their classrooms is a complicated matter. Finding ways to support and nurture, while maintaining high expectations in a virtual environment is daunting. It requires individualization and personal touch that is feasible when $N=10$. Sequencing topics, amassing appropriate curricular units and readings for nourishment and exploration, building and sustaining productive F2F and virtual communities of practice with teachers representing urban, rural, suburban, middle and high school mathematics and science contexts is a complex endeavor and requires a well-stocked arsenal of resources.

Teachers experienced shared activities during F2F sessions that challenged them to make sense of statistical concepts, with and without technology, as well as provide pedagogical modeling to consider. These sessions developed a sense of community and fostered relationships that promoted productive virtual collaboration. Because the virtual modules and curriculum projects allowed students to "choose their own adventure," they could target concepts and resources most relevant to their work or interests. This autonomy appeared welcome and novel for teachers.

Ten teachers completed eight statistical curriculum implementation projects requiring them to reflect on their students' learning. Six students worked independently and four students partnered up. Every project incorporated dynamic statistical technology, some multiple tools. Each project demonstrated student learning through collected artifacts including classroom video and student work samples. Given the written documentation of the plans, descriptions of the implementation, and reflections on the unit with at least one peer, it is clear that all of these teachers extended their TPSK. Their enactments were informed by literature and course experiences. They often referred to *Core-Plus Mathematics* (Hirsch et al. 2015) units and the *GAISE* (Franklin et al. 2007) document for guidance and courageously went live with real students with new and challenging content while utilizing and helping their students use new tools. Evidence in the form of self-assessments, instructor assessment, and participants' written artifacts suggests that the *doer to designer* inspired blended course design with summer/fall timeline has been impactful for teachers' personal learning of statistics and modeling relevant for the secondary curriculum, thus improving their TPSK. Furthermore, there is mounting evidence that teachers' thinking about statistical instruction has evolved toward a more sense-making, activity-based, technology-oriented perspective, suggesting the approach is promising.

Acknowledgements This material is based upon work supported by the National Science Foundation under Grant No. 1136392. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Australian Curriculum Assessment and Reporting Authority. (2010). *Australian curriculum consultation portal*. Retrieved from <http://www.australiancurriculum.edu.au/Home>.
- Batanero, C., Burrill, G., & Reading, C. (2011). *Teaching statistics in school mathematics—Challenges for teaching and teacher education: A joint ICMI/IASE study: the 18th ICMI study*. Dordrecht: Springer.
- College Board. (2015). *Advanced placement biology: Course and exam description*. Retrieved from <https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap-biology-course-and-exam-description.pdf>.
- Conference Board of the Mathematical Sciences. (2010). *The mathematical education of teachers II* (Vol. 17). Providence, RI: American Mathematical Society.
- Fendel, D., Resek, D., Alper, L., & Fraser, S. (2012). *Interactive mathematics program*. Emeryville, CA: Key Curriculum Press.
- Finzer, W. (2005). *Fathom 2™ dynamic data software*. Emeryville, CA: Key Curriculum Press.
- Franklin, C., Kader, G., Bargagliotti, A., Scheaffer, R., Case, C., & Spangler, D. (2015). *Statistical education of teachers*. Retrieved from American Statistical Association: <http://www.amstat.org/asa/files/pdfs/EDU-SET.pdf>.
- 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.
- Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, 38(5), 427–437.
- Hirsch, C. R., Fey, J. T., Hart, E. W., Schoen, H. L., & Watkins, A. E. (2015). *Core-plus mathematics: Contemporary mathematics in context*. Columbus, OH: McGraw-Hill Education.
- Kadjevich, D. M., & Madden, S. R. (2015). Comparing approaches for developing TPCK. In *New directions in technological and pedagogical content knowledge research: Multiple perspectives* (pp. 125–146): Information Age Publishing.
- Keller, S. (2006). *CPMP-tools*. East Lansing, MI: Michigan State University and Core-plus mathematics project.
- Konold, C., & Miller, C. (2005). *Tinkerplots™ dynamic data exploration*. Emeryville, CA: Key Curriculum Press.
- Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S., & Phillips, B. (2009). *Connected mathematics 2*. Boston, MA: Pearson.
- Lee, H. S., & Hollebrands, K. (2011). Characterizing 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). Springer.
- 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 Paper presented at the Sustainability in Statistics Education. Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, AZ.
- Madden, S. R. (2008). High school mathematics teachers' evolving understanding of comparing distributions. *Dissertation*. Kalamazoo, MI: Western Michigan University.
- Meyer, O., & Lovett, M. (2014). *Using Carnegie Mellon's open learning initiative (ODI) to support the teaching of introductory statistics: Experiences, assessments, and lessons learned*. Paper presented at the International Conference on Teaching Statistics 9, Flagstaff, AZ.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 180(6), 1017–1054.
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*. Washington, DC: Authors.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: The National Council of Teachers of Mathematics.

- National Council of Teachers of Mathematics. (2009). *Focus in high school mathematics: Reasoning and sense making*. Reston, VA: The National Council of Teachers of Mathematics Inc.
- National Research Council (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/18290>.
- Papert, S. (1991). Situating constructionism. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 1–12). Norwood, NJ: Ablex Publishing Corporation.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In J. F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 957–1010). Charlotte, NC: Information Age Publishing.
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK: Cambridge University Press.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67, 223–265.

Chapter 13

Statistical Reasoning When Comparing Groups with Software—Frameworks and Their Application to Qualitative Video Data



Daniel Frischemeier

Abstract Comparing groups is a fundamental activity in statistics. Preferably such an activity is embedded in a data analysis cycle and done with real and large datasets. Software enables learners to carve out many differences between the compared distributions. One important aspect in statistics education is how to evaluate these complex intertwined processes of statistical reasoning and the use of software when comparing groups. The primary intention of this chapter is to introduce a framework for evaluating statistical reasoning and software skills when comparing groups and to show an application of this framework to qualitative data collected during a video study of four pairs of preservice teachers comparing groups with TinkerPlots.

Keywords Frameworks · Group comparisons · Qualitative content analysis
Statistical reasoning · TinkerPlots

13.1 Introduction

Statistics has received more and more attention in school mathematics in the last 15 years in Germany. The leading idea “Data, Frequency and Chance” (Hasemann and Mirwald 2012) recommends the implementation of statistics beginning at the primary school level. In grades 5–10 in secondary schools in Germany the leading idea “Data & Chance” (Blum et al. 2006) demands—amongst others—the implementation of a data analysis cycle (similar to the PPDAC cycle, which consists of the phases Problem, Plan, Data, Analysis and Conclusions, see Wild and Pfannkuch 1999). In addition the work with real and rich data and the use of adequate software is recommended. Since data explorations in large data sets are inevitably connected with competent software use, preservice teachers need a good background not only in statistical content but also in technological knowledge (see Lee and Hollebrands 2011). These facts set requirements not only for schools and teachers but also for

D. Frischemeier (✉)
Paderborn University, Paderborn, Germany
e-mail: dafr@math.upb.de

© Springer Nature Switzerland AG 2019
G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_13

universities who are responsible for the education of upcoming teachers in statistics. Recommendations for elementary preservice teachers' education in statistics can be found in the recommendations of Arbeitskreis Stochastik (2012) on the German national level and in Batanero et al. (2011) on an international level. Crucial aspects, which can be found in both recommendations, are engagement in a data analysis cycle (like PPDAC), exploration of real data with adequate software and interpretation of findings. The comparison of distributions, which is, according to Konold and Higgins (2003, p. 206) "the heart of statistics", can be seen as a fundamental activity in statistics since this activity includes many of the fundamental ideas raised by Burrill and Biehler (2011) such as data, variation, distribution and representation. When comparing groups in real and rich datasets the use of software is important for an explorative data analysis and for switching between different displays. The software TinkerPlots (Konold and Miller 2011) offers several features, which can help learners to compare groups and can be considered as adequate educational software to learn data analysis and as a powerful tool for exploring data. More explicitly TinkerPlots can be used for different purposes: as educational software for pupils from grade 4 to 10, as a data analysis tool for preservice teachers and as a medium for teachers to demonstrate data analysis in classroom. This chapter presents a framework to evaluate statistical reasoning when comparing groups with software (TinkerPlots) on the basis of existing research (see the literature review below) and personal experiences. The framework is used to analyze qualitative data in the form of transcripts created from the recorded video data. The video study, which is itself embedded in a larger project (Frischemeier 2017), had the aim to investigate how preservice teachers compare groups in a real and rich dataset with TinkerPlots after a course for preservice teachers on data analysis with TinkerPlots at Paderborn University.

13.2 Literature Review

The following provides a short overview of different trends in the research literature on learners' reasoning when comparing groups. The literature review took into account research articles and studies dealing with learners' strategies for group comparisons and the use of frameworks to evaluate the strategies. One major aim was to generate a framework for a video study to evaluate the group comparison skills of a sample of preservice teachers with TinkerPlots, e.g. how capable are the preservice teachers in comparing groups with TinkerPlots after the course.

A first trend in the research literature can be seen using the SOLO taxonomy (Biggs and Collis 1982) for learners' outcomes when comparing groups. Watson and Moritz (1999) rated group comparison skills of Australian 3–8 graders via a SOLO taxonomy in the levels "unistructural", "multistructural" and "relational". They conducted interviews where the participants were asked to compare distributions of test scores of school classes ("which class is better?") in different settings. All distributions were given as stacked dot plots. Three of the four interview tasks included a group comparison where both groups had equal sizes. In the fourth group comparison task,

the setting was non-equal-sized. In the interview learners used numerical and visual strategies to compare groups and to answer the question, which class was better. Watson and Moritz (1999) found out—amongst other things—that more students in higher grades tend to reason proportionally than do students in lower grades.

In a second trend, a focus on the concept of variability when comparing groups can be identified. Makar and Confrey (2002) concentrated on preservice teachers doing group comparison tasks with Fathom and generated a “taxonomy for classifying levels of reasoning when comparing two groups”. Here they focused on the way the learners used inferential reasoning to compare—amongst other things—variability between two groups. In an interview study with preservice teachers Makar and Confrey (2004), as a subsequent study to Makar and Confrey (2002), identified three different aspects of learners’ reasoning about variability when comparing two groups: “as variation within a group—the variability of data”, “as variability between groups—the variability of measures” and “distinguishing between these two types of variability” (see Makar and Confrey 2004, p. 368). One major implication from their research is that learners seem to be capable of identifying variability within groups but have difficulties when comparing groups with regard to variability. Learners taking into account variability when comparing groups were also the focus of a study by Ben-Zvi (2004), in which he observed that learners working on a group comparison task at first concentrate on the variability of the distributions and later take into account differences in regard to center, shape and outliers in the data.

Biehler (2001) postulated a normative view of group comparisons, which he views as adequate strategies for doing group comparisons. According to Biehler (2001) comparisons of two distributions of numerical variables are called *p*-based, if, for some x the relative frequencies $h(V \geq x)$ and $h(W \geq x)$ are compared; in *p*-based comparisons a specific argument x can be given (for example: 10 h), and the proportion of cases that are equal or larger than 10 h is compared in both groups (Biehler 2001, p. 110). In addition, comparisons of two distributions of numerical variables are called *q*-based, if, for some proportion p between 0 and 1, the matching quantiles of the variables V and W , $q_V(p)$ with $q_W(p)$, are compared. With $q(p)$ is the quantile with respect to p . For $p = 0.5$ this is a comparison of medians (Biehler 2001, p. 110). Biehler (2007) points out further ideas and elements when comparing distributions such as comparing the skewness of the distributions or using the so called “shift model” to identify an additive or a multiplicative shift between the distributions (for details see Biehler 2007).

Pfannkuch (2007), in a research study preceding Pfannkuch et al. (2004), developed a framework for rating learners’ skills comparing boxplot distributions, which can also be used for rating learners’ group comparison skills in general. In this framework Pfannkuch (2007) distinguishes several comparison elements such as summary, signal, spread and shift, at first. Then Pfannkuch (2007, p. 159) rates the quality of the comparison of each element in different hierarchical quality levels: “point decoder” (level 0), “shape comparison describer” (level 1), “shape comparison decoder” (level 2) and “shape comparison assessor” (level 3). Whereas a “point decoder” identifies differences of single values of the distributions and a “shape comparison describer” makes comparison statements on a descriptive level, the highest level, the “shape

comparison assessor” interprets differences between the distributions. The “shape comparison decoder” can be seen as an intermediate step between level 1 and level 3 (for further details see Pfannkuch 2007, p. 159). This framework was also used by Pfannkuch (2007) to analyze the outcomes of an empirical study where learners were asked to compare distributions (given in the form of boxplots). Pfannkuch (2007) observed that the participants preferred to compare the distributions via summary and spread rather than via shift or signal elements. In addition, while most participants worked out differences in the “describing” or “decoding” level, only a few worked out differences in the highest level (“assessor”).

A fifth trend in research literature can be identified in regard to software use when comparing groups. Biehler (1997) emphasized the role of software when doing data explorations and when comparing groups and identified four phases: “statistical problem”, “problem for the software”, “results of software use” and “interpretation of results in statistics” (Biehler 1997, p. 175). Frischemeier (2014) has refined the phases inductively (taking into account the findings from the data analysis) and distinguished between six phases when comparing groups with software (“real problem”, “statistical activity”, “software use”, “reading off/documentation”, “conclusions” and “reasons”). Biehler (1997) and Frischemeier (2014) noted that learners in their work tended to jump into a statistical problem without considering the real problem and that the learners produced displays with the software but do not interpret them with regard to the statistical problem. There is scarce research with regard to specific software skills of learners, when using software for their group comparison tasks. Here the work of Maxara (2009, 2014), in which a framework for skills using Fathom for simulating chance experiments is given, can be adapted to a framework for more general skills for digital tools (like TinkerPlots).

From the five different trends and frameworks presented in the literature review, two main dimensions can be derived:

- D1: Statistical reasoning when comparing groups (Which elements are compared by the learners? How do learners interpret their worked out differences?)
- D2: Skills of using software (e.g. TinkerPlots) when comparing groups (In which way are learners capable of using the software in a competent way for their purposes?)

For the component “Skills of using software (TinkerPlots) when comparing groups”, the framework from Maxara (2009, 2014) about Fathom competences when simulating chance experiments with Fathom will be adapted for the purposes of this study. Accordingly for the dimension “Statistical reasoning when comparing groups”, six elements for comparing groups: center, spread, skewness, shift, p -based comparisons, and q -based comparisons are identified from the framework of Pfannkuch (2007) and from the work of Biehler (2001, 2007). These elements are taken as sustainable for group comparisons, and preservice teachers should carve out differences between groups regarding them (as in Pfannkuch 2007; see below): “point decoder” (level 0), “shape comparison describer” (level 1), “shape comparison decoder” (level 2) and “shape comparison assessor” (level 3).

As mentioned above the main intention of this chapter is to create frameworks to evaluate the statistical reasoning of preservice teachers when comparing groups with software, specifically TinkerPlots. In the following the frameworks used in the study are derived from the literature in regard to the two dimensions D1 and D2. The frameworks are then applied to the analysis of qualitative data collected in a video study where preservice teachers have compared groups with TinkerPlots using real datasets.

13.3 Frameworks for Evaluating Statistical Reasoning When Comparing Groups with Software

13.3.1 *Framework for D1 “Statistical Reasoning When Comparing Groups”*

One intention with regard to the dimension “Statistical reasoning when comparing groups” was (similar to Pfannkuch 2007) to identify the group comparison elements used by the study participants first and then rate the quality of the comparisons. The categories (structural elements) for this framework “center”, “spread”, “shift”, “skewness”, “*p*-based” and “*q*-based” have primarily arisen deductively on basis of the literature review of Pfannkuch (2007) and Biehler (2001, 2007). From the elements “signal”, “shift”, “summary” and “spread” of Pfannkuch (2007) the comparison element of “summary” was specified as a comparison element of “center”, the comparison element “signal” was left out since this kind of comparison element seemed to be too specific (for boxplots), so finally the comparison elements “center”, “spread” and “shift” were identified. Based on Biehler (2001, 2007) the comparison elements “skewness”, “*p*-based” and “*q*-based” were adopted. In a further step the hierarchical levels of Pfannkuch (2007), “point decoder” (level 0), “shape comparison describer” (level 1), “shape comparison decoder” (level 2) and “shape comparison assessor” (level 3), were merged, and new hierarchical levels “low quality”, “medium quality” and “high quality” were defined for each structural element (center, spread, shift, skewness, *p*-based, *q*-based). This lays the foundation for a high quality comparison related to the elements mentioned above where the elements are compared in a quantitative way and are interpreted. For example, a high quality comparison statement about the difference in income between men and women regarding “center” would be “men earn 29.5% more than women on average”, because the difference is quantified (“29.5%”) and interpreted (“men earn ... more than women on average”). When group comparison elements are only compared in a qualitative way but not interpreted, they are rated as medium quality (example: “The mean of men is higher than the mean of women”), because in this case it is only said that the mean of distribution X is larger than the mean of distribution Y. The case where the measures or elements are compared in a wrong way is called “low quality”. Table 13.1, adapted

from Frischemeier and Biehler (2016, p. 646), displays the definitions of the group comparison elements in different quality levels.

13.3.2 Framework for D2 “Skills of Using Software (TinkerPlots) When Comparing Groups”

The categories with regard to TinkerPlots skills take into account the categories of Maxara (2009, 2014) about Fathom competencies when simulating chance exper-

Table 13.1 Group comparison elements and different quality levels

| | High quality | Medium quality | Low quality |
|-----------------|---|--|---|
| Center | Measures of center (mean, median) are compared in a quantitative way and are interpreted. | Measures of center (mean, median) are compared in a qualitative way but are not interpreted. | Measures of center (mean, median) are compared incorrectly. |
| Spread | Measures of spread (IQR) or informal descriptions of spread (such as “density”, “close”) are compared and interpreted. | Measures of spread (IQR) or informal descriptions of spread (such as “density”, “close”) are compared but not interpreted. | Spread is compared with inadequate measures (like range) and/or is interpreted incorrectly. |
| Shift | Shift between both distributions is quantified correctly (comparing the position of the middle 50% or using the “shift model”). | Shift between both distributions is described in a qualitative way (e.g. comparing non-corresponding numbers) | Shift between both distributions is worked out incorrectly. |
| Skewness | Skewness of both distributions is described correctly, and the differences between the distributions are interpreted. | Skewness of both distributions is described correctly, but the differences are not interpreted. | Differences of skewness are worked out incorrectly. |
| <i>p</i> -based | <i>p</i> -based differences are identified and interpreted. | <i>p</i> -based differences are identified but not interpreted. | <i>p</i> -based differences are worked out incorrectly. |
| <i>q</i> -based | <i>q</i> -based differences are identified and interpreted. | <i>q</i> -based differences are identified but not interpreted. | <i>q</i> -based differences are worked out incorrectly. |

Key examples related to spread, skewness, shift, *p*-based and *q*-based comparisons can be read in Frischemeier and Biehler (2016, p. 647)

Table 13.2 Definitions and examples related to “TinkerPlots skills” when comparing groups

| TinkerPlots skill | Definition | Example |
|--------------------------|---|--|
| TinkerPlots skill high | Learners have a concrete plan in mind and can fulfill it with TinkerPlots. | Conrad & Maria: “Let’s do a boxplot”. Conrad and Maria produce a boxplot in TinkerPlots. |
| TinkerPlots skill medium | Learners have a concrete plan in mind and can fulfill it with TinkerPlots after a trial-and-error approach. | Hilde & Iris are unsure which button is for displaying the mean and which button is for displaying the median. |
| TinkerPlots skill low | Learners have a concrete plan in mind and cannot fulfill it with TinkerPlots. | Laura & Ricarda want to plot a boxplot in TinkerPlots. After some time Laura said: “I do not know how”. |

iments with Fathom. We have refined them in our data analysis process. A high TinkerPlots skill indicates that learners want to perform a certain action with TinkerPlots and can use TinkerPlots for their purpose. To rate TinkerPlots skills in this way, participants in the video study were asked before every action with TinkerPlots to articulate and describe their plan and intention. As an example consider the pair Conrad and Maria (discussed further later in the section about the video study). Conrad and Maria want to display a boxplot of the distribution of the variable “monthly income” and articulate: “let’s do a boxplot”. They immediately used TinkerPlots to produce a boxplot of the distribution of their desired variable. This indicates a “TinkerPlots skill high”. Another example of a high TinkerPlots skill can be seen in Fig. 13.2. Here Hilde and Iris (again revisited later in the section about the video study) discuss how to find the ratio of low-wage earners in both distributions (see transcript excerpt, Fig. 13.2, line 01-03), and they successfully use dividers in TinkerPlots (see transcript excerpt, Fig. 13.2, line 04-06) to figure out the desired ratios. If the participants in a situation similar to the one above have a certain intention with TinkerPlots but can only realize this intention via a “trial-and-error”-approach this is “TinkerPlots skill medium”. As an example consider again the working phase of Hilde and Iris: Hilde and Iris are unsure which button they have to use to display the mean and which button they have to use to display the median of the distribution of monthly income. In the case where the learners have a concrete intention with TinkerPlots but cannot realize it with the software, the situation is rated as “TinkerPlots skill low”. An example here is given in the working process of Ricarda and Laura (also discussed further in the section about the video study). Laura and Ricarda want to plot a boxplot in TinkerPlots, but after some time Laura says: “I do not know how” and they do not plot a boxplot in TinkerPlots. In Table 13.2 we see the overview for all definitions and key examples related to “TinkerPlots skills” when comparing groups.

13.4 Application of the Framework to Qualitative Data Collected in a Video Study on Preservice Teachers Comparing Groups with TinkerPlots

The focus of this study was on how preservice teachers compare groups with TinkerPlots after attending a course on data analysis with TinkerPlots called “Developing statistical thinking and reasoning with TinkerPlots”.

The following summarizes the contents of the course and then considers how the framework was applied. The choice of the task used in the video study, the participants, the data collection and the data analysis process of the video study are described. Finally the framework is applied to the video data and the written work of four pairs of preservice teachers working on group comparison tasks with TinkerPlots.

13.4.1 The Course “Developing Statistical Thinking and Reasoning with TinkerPlots”

The course “Developing statistical thinking and reasoning with TinkerPlots” was designed using the Design Based Research paradigm (Cobb et al. 2003). Fundamental ideas for this course are to “focus on developing central statistical ideas” , to “use real and motivating data sets” , to “use classroom activities to support the development of students’ reasoning”, the integration of “appropriate technological tools” , to “promote classroom discourse that includes statistical arguments and sustained exchanges that focus on significant statistical ideas” and the “use of formative assessment” (Garfield and Ben-Zvi 2008, p. 48). An overarching structure was the PPDAC-cycle (Wild and Pfannkuch 1999). The participants were supposed to work with real and multivariate data that they collected themselves (via questionnaire or via downloading official data files from the internet), explored them with regard to self-generated statistical questions using TinkerPlots, and then wrote down their findings in a statistical report. All in all, the course consisted of four modules. In the first module the participants got to know the PPD (Problem, Plan, Data) elements of the PPDAC-cycle. Here the participants learned how to generate adequate statistical (research) questions, to plan a data collection, to generate questionnaires, to collect data, to conduct a data collection, to store data and to clean data on their own. Module two included an introduction to data analysis with TinkerPlots. Here the participants were shown how to produce plots for univariate and bivariate data explorations using TinkerPlots and how to describe and interpret them with regard to fundamental elements like center, variation, skewness, etc. (Rossman et al. 2001; Biehler 2007). One main idea was to develop statistical reasoning in conjunction with cooperative learning environments like “think-pair-share” (Roseth et al. 2008). At first the participants worked on a data exploration on their own (“think”), then discussed it with partners (“pair”) and finally with experts (the teachers) in plenum

("share"). The peer feedback by the partners and the expert feedback by the teachers was intended to develop the quality of the statistical investigations in module two and to prepare them for wider explorations in module three. In module three, a large part of the course was dedicated to wider explorations on differences among two or more groups concentrating on the comparison of groups regarding the elements center, spread, shift, skewness, p -based and q -based (see Pfannkuch 2007; Biehler 2007; Frischemeier 2017), which were also taken into account for the framework. There were two norms set for the participants with regard to group comparisons: first, to work out as many differences on the six elements as possible; second, to interpret these differences, which are carved out between the two groups regarding center, spread, shift, skewness, p -based and q -based comparisons. Boxplots were highlighted as adequate displays for group comparisons; the use of the "shift model" was highlighted for identifying possible additive or multiplicative shifts between two distributions. Also the participants were shown how to create different group comparison displays with TinkerPlots like histograms, hatplots, boxplots and stacked dot plots. In addition the participants were shown how to use dividers for p - and q -based comparisons in TinkerPlots. Finally, in the fourth module the participants were introduced to inferential statistics in the sense of conducting randomization tests with TinkerPlots (Frischemeier and Biehler 2014) as a succeeding activity to comparing groups. Further details, lessons plans and materials used in the course can be found in Frischemeier (2017).

13.4.2 Task of the Video Study

The task used in this video study included a group comparison activity in a real, large and motivating dataset, which was a random sample from a large and multivariate dataset (called VSE data) imported from the German Bureau of Statistics containing data on 861 persons from the German population and variables such as salary, gender and region. Six weeks after the course "Developing statistical thinking and reasoning with TinkerPlots", eight participants from the course were invited to take part in a video study, in which they were asked to compare groups in the VSE data with TinkerPlots. The precise task they were to work on was: *In which aspects do the distributions of the variable "salary" of the men and the women differ? Carve out as many differences in both distributions as you can.* There are plenty of possibilities to carve out differences between both groups with TinkerPlots. Figure 13.1 shows different representations learners might use in TinkerPlots to work out differences between both distributions regarding center, spread, shift and q -based comparisons (Fig. 13.1, left), with regard to skewness (Fig. 13.1, centered) and with regard to p -based comparisons (Fig. 13.1, right). For details and possible ways to solve the VSE task see Biehler and Frischemeier (2015). From this initial point of possible ways to carve out differences between both distributions and from the two dimensions "Statistical reasoning when comparing groups" and "Skills of using software (TinkerPlots) when comparing groups" the main research question was: *How do the*

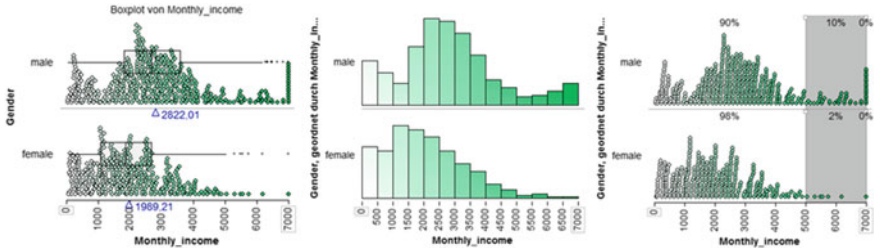


Fig. 13.1 Possible TinkerPlots graphs for VSE task, created with the German version of TinkerPlots

preservice teachers in the study compare groups in this real dataset with TinkerPlots after the course? More precisely three sub-questions can be distinguished for the analysis of the video study:

- Which comparison elements do the preservice teachers use when comparing groups in a real dataset in TinkerPlots?
- Do they interpret the differences they worked out for the comparisons?
- In which ways are the preservice teachers able to use the software TinkerPlots for their purposes when comparing groups?

13.4.3 Data and Participants

The eight participants (all participants attended the course “Developing statistical thinking and reasoning with TinkerPlots” described above) worked on the task in a laboratory setting at the Paderborn University in groups of two. They were given an exercise sheet with the task, a list of variables of the dataset and also the TinkerPlots file containing the VSE data. Furthermore they were told to communicate aloud as much as possible during their working process. There was no intervention at any time in their working process by the interviewer. Their screen activities, their oral communication, and the action with TinkerPlots were recorded, and also their notes on their exercise sheets were collected. The communication between the participants and action with the software was transcribed taking into account the recommendations of Kuckartz (2012). As an example see the following abstract from a transcript of Hilde (H) and Iris (I) in Fig. 13.2. The software actions are documented in parentheses; each TinkerPlots action is also documented with a screenshot. Communication between both participants is displayed in normal wording.

| | | |
|----|----|--|
| 01 | H: | Let us have a look on the low-wage employees. This might be interesting. #00:11:44# |
| 02 | I: | Hmhm. #00:11:44# |
| 03 | H: | Figure out, how many are below or equal 1000 € with the divider. #00:11:50# |
| 04 | H: | <p>Boxplot von Monthly_income</p> <p>female: 0% 22% 1080 1860 2680 3580 78%</p> <p>male: 0% 14% 1080 1860 2680 3580 86%</p> <p>Monthly_income</p> <p>(Both dividers are adjusted on 1000€)</p> |
| 05 | I: | Female employees: 22%/ #00:12:25# |
| 06 | H: | Male employees: 14% #00:12:49# |

Fig. 13.2 Excerpt of transcript of Hilde (H) and Iris (I) when working on VSE task

13.4.4 Data Analysis

The analysis of data was done with a qualitative content analysis approach (Mayring 2015). The scheme and the cycle of a qualitative content analysis can be seen in Fig. 13.3.

When applying the qualitative content analysis, the first step is to define the object of research and the research question. For the video study reported in this section, the research question was: “How do our preservice teachers compare groups in this real dataset with TinkerPlots after the course?”. In a second step the dimensions of analysis are derived from the research question and the object of research (and also from the existing field of literature). In this case two dimensions were identified—as it was already done above after the literature review: “Statistical reasoning when comparing groups” (D1) and “Skills of using software (TinkerPlots) when comparing groups” (D2). Then (step 3) units of analysis (which data are taken into account for the analysis?), minimum coding units (what is the smallest piece in the data taken into account for analysis?) and maximum coding units (what is the largest piece in the data taken into account for analysis?) are defined. In step 4 categories with exact definitions and key examples are defined and coding schemes (see frameworks in Tables 13.1 and 13.2) are set up. Then selected cases from the data material are coded

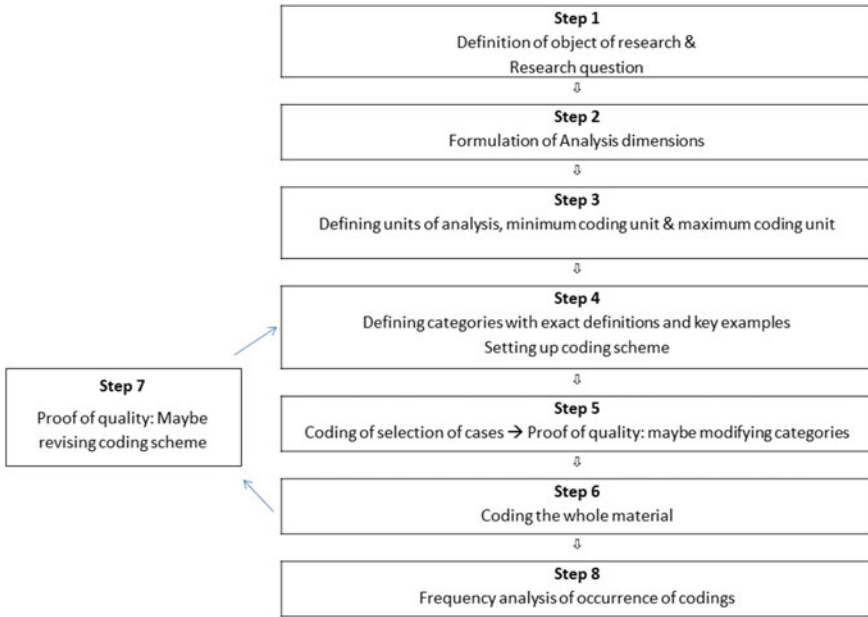


Fig. 13.3 Scheme of qualitative content analysis in this study

followed by a first proof of quality by an independent researcher (step 5). If there are many discussion points with the independent researcher, the categories have to be modified and the whole material has to be coded with the modified categories again (step 6). After step 7 the main proof of quality is implemented. The frameworks with its definitions and key examples are given to an independent researcher. The independent researcher codes the transcripts on the base of the frameworks. The elements where the codings of the independent researcher are different from the codings of the researcher will be discussed, and the intercoderreliability will be calculated on the base of the formula below (see Mayring 2010, p. 120). Here x means the number of codes that match between researcher and independent researcher, n means the whole number of codes and k means the number of categories. Mayring (2001) says, that $\kappa \geq 0.7$ is adequate for proving intercoderreliability. In the case of an adequate intercoderreliability, a frequency analysis of occurrence of codings is done in step 8.

$$\kappa = \frac{\frac{x}{n} - \frac{1}{k}}{1 - \frac{1}{k}}$$

If the intercoderreliability is inadequate, the coding scheme/framework has to be revised—the procedure will then be continued with step 4.

One crucial point (see Mayring 2015, p. 374) in qualitative content analysis, which can be seen in Fig. 13.3 (step 4), is a coding scheme/framework with exact defini-

tions and key examples. A dynamic framework (see Table 13.1) was used to identify the variety of group comparison elements and their quality. In the analysis of the comparison elements and their quality, the focus was on the transcribed communication of the working phase and on the written notes of the participants collected during the video study. Therefore, the analysis unit for dimension D1 covers the transcribed conversation in-between the pairs and their written notes. Furthermore a framework was created to rate and evaluate the TinkerPlots skills of the participants (see Table 13.2). For coding the “TinkerPlots skills” the analysis unit covers the transcribed conversation in-between the pairs and their action with the software. For both frameworks, a word was chosen as minimum coding unit, and a unit of meaning was chosen as maximum coding unit. The coding units were assigned to the codes in a disjointed way. Intercoderreliability was established by comparing the codings with an independent researcher ($\kappa = 1.000$ for dimension D1 and $\kappa = 0.8875$ for dimension D2).

The following sections present the application of the frameworks to evaluate statistical reasoning and software use when comparing groups.

13.4.5 Results for “Statistical Reasoning When Comparing Groups” (D1)

A broad range of 28 comparison elements was used by the preservice teachers (see Table 13.3) when working on the VSE task. The most used elements were p -based comparisons (eight times) and comparisons via shift (six times). Comparisons with regard to skewness and q -based comparisons were the least used. There was no comparison rated with a low quality. 28.6% of the comparisons were done on a high quality level, 71.4% on a medium quality level. Remarkably all p -based and shift comparisons have been done in medium quality, so there might be a difficulty for the preservice teachers to interpret these specific differences. With regard to the comparisons via shift, no one used the “shift model”, although it was part of the curriculum and it appeared in the data. Table 13.3 shows the distribution of codings regarding group comparison elements and their quality used by the four pairs when working on the VSE task.

The following applies the framework with respect to “Statistical reasoning when comparing groups” to the work of the four pairs.

Conrad and Maria and Sandra and Luzie made only four comparison statements each. All of them are rated with medium quality. Conrad and Maria compared the distributions with regard to spread and shift and also used a q -based comparison. Amongst other graphs, they produce a stacked dot plot with boxplots of the variable monthly income separated by gender in TinkerPlots (Fig. 13.4).

For their q -based comparison they read off the minimum and the 1st quartile of each distribution in TinkerPlots (with the help of the mouse cursor) and documented their findings (see their written documentation in Fig. 13.5).

Table 13.3 Frequency analysis of codings related to the group comparison elements and their quality

| | High quality | Medium quality | Low quality | Overall |
|-----------------|--------------|----------------|-------------|-------------|
| Center | 2 | 2 | 0 | 4 |
| Spread | 3 | 2 | 0 | 5 |
| Shift | 0 | 6 | 0 | 6 |
| Skewness | 2 | 1 | 0 | 3 |
| <i>p</i> -based | 0 | 8 | 0 | 8 |
| <i>q</i> -based | 1 | 1 | 0 | 2 |
| Overall | 8 (28.6%) | 20 (71.4%) | 0 (0.0%) | 28 (100.0%) |

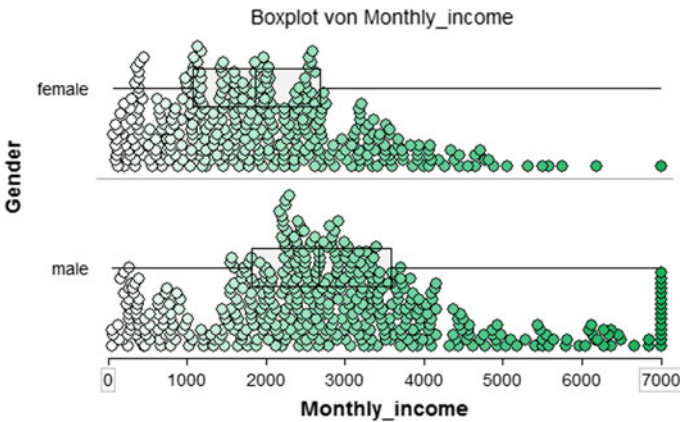


Fig. 13.4 TinkerPlots graph produced by Conrad and Maria

The lower 25% of the women earn between 71€ and 1076,5€.
 The lower 25% of the men earn between 47€ and 1825€.

Fig. 13.5 Translated *q*-based comparison done by Conrad and Maria (written note)

This *q*-based comparison as can be seen in Fig. 13.5, was rated with a medium quality since the intervals of the lower 25% are just described but there is no comparison done between both groups.

Sandra and Luzie—like Conrad and Maria—also only made four comparison statements. Remarkably they only used *p*-based comparisons to work out differences between the distributions. All of them were rated with medium quality, since none of the differences were interpreted nor compared directly. Figure 13.6 displays Sandra

and Luzie’s TinkerPlots graph for a p -based comparison where they identified the relative frequencies of men and women having an income larger than 5000€.

Their findings can be seen in the documentation in Fig. 13.7. Since the relative frequencies are documented, but the differences are neither interpreted nor compared, we rate this comparison as a “medium quality p -based comparison”.

Hilde and Iris and Laura and Ricarda did ten comparisons each and used a large range of comparison elements. Also high quality comparisons can be found in their communication and working process.

Hilde and Iris worked out differences between both distributions via center, spread, shift and p -based comparisons. The differences regarding center, shift and p -based comparisons were worked out in medium quality. For an example of one of their p -based comparisons of medium quality see the transcript excerpt in Fig. 13.2. Here Hilde and Iris figured out the relative frequency of employees earning 1000€ or less per month (their definition of low-wage employees). They concluded that 22% of the female employees and 14% of the male employees are low-wage employees. To work out differences in regard to spread, Hilde and Iris used stacked dotplots and boxplots in TinkerPlots (see Fig. 13.8). Furthermore they used so-called reference lines in TinkerPlots to determine summary statistics like median, 1st quartile and 3rd quartile of both distributions.

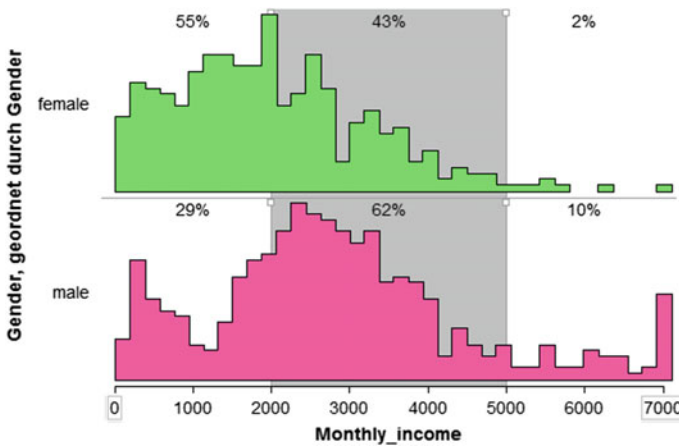


Fig. 13.6 TinkerPlots graph produced by Sandra and Luzie

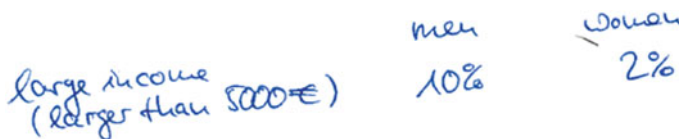


Fig. 13.7 Translated p -based comparison done by Sandra and Luzie (written note)

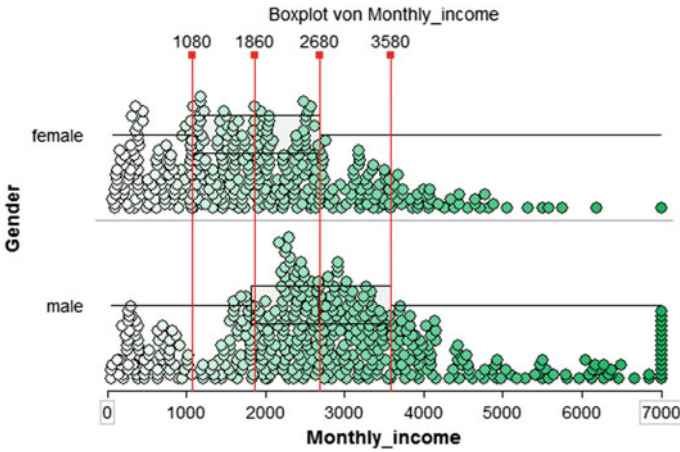


Fig. 13.8 TinkerPlots graph produced by Hilde and Iris

Both differences related to spread used by Hilde and Iris were rated as “high quality comparison”. One example can be seen in their written note in Fig. 13.9. Here Hilde and Iris read off the 1st and the 3rd quartiles of both distributions and calculated the interquartile range (IQR) of each distribution. Their conclusion then is that “the middle 50% of men spreads more than the middle 50% of the women”, which was rated as a high quality comparison since the differences in regard to spread are worked out correctly and are interpreted (“spreads more”).

All in all two of their ten comparisons (20%) are rated as high quality.

Laura & Ricarda showed the best performance of all pairs: They did ten comparisons and worked out differences between both distributions with regard to all six group comparison elements. Six of their ten comparisons (60%) were rated as “high quality”. In one of their comparisons in regard to center Laura and Ricarda

Fig. 13.9 Comparison of spread done by Hilde and Iris (written note)

WOMEN : $Q_1 = 1090$; $Q_3 = 2650$
 IQR = 1600
 MEN : $Q_1 = 1820$; $Q_3 = 3600$
 IQR = 1800
 → THE MIDDLE 50% OF THE MEN SPREADS MORE THAN THE MIDDLE 50% OF THE WOMEN.

first displayed both distributions as horizontally stacked dotplots in TinkerPlots and then calculated the mean of the distributions of “monthly income” (Fig. 13.10).

As one of their two comparisons related to center, Laura and Ricarda stated that “in 2006 the men earn 29.5% more on average than the women” (see their written documentation in Fig. 13.11).

This comparison was rated as high quality because it quantifies the differences between the means in a multiplicative way (29.5%) and because this difference is interpreted (“on average”).

13.4.6 Results for “Skills of Using Software (TinkerPlots) When Comparing Groups”(D2)

Table 13.4 shows the distribution of codings related to “Skills of using software (TinkerPlots) when comparing groups”. Examples can be seen in the description of the framework D2 above.

The data suggests that TinkerPlots is used in a very competent way by the participants in the study. Approximately three quarters (75.9%) of all codings are related to high TinkerPlots skills. About 86% of the codings are related to at least medium

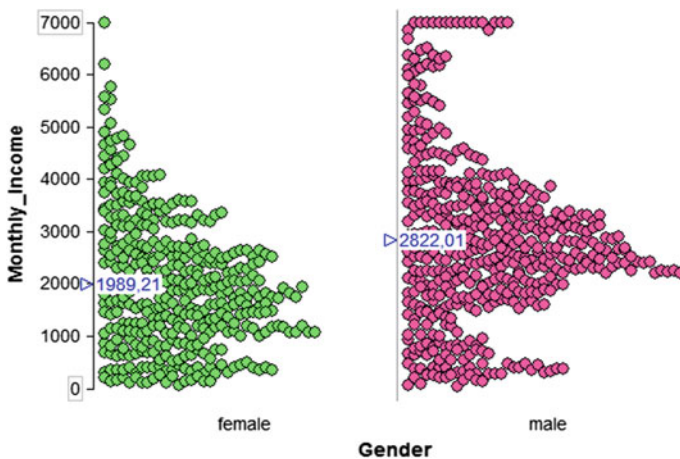


Fig. 13.10 TinkerPlots graph produced by Laura and Ricarda

IN 2006 THE MEN EARN 29.5% MORE ON AVERAGE THAN THE WOMEN.

Fig. 13.11 Comparison of center done by Laura and Ricarda (written note)

Table 13.4 Frequency analysis of codings related to the TinkerPlots skills by all pairs

| | TinkerPlots skill high | TinkerPlots skill medium | TinkerPlots skill low | Overall |
|-----------------|------------------------|--------------------------|-----------------------|--------------|
| Conrad & Maria | 22 (75.9%) | 1 (3.4%) | 6 (20.7%) | 29 (100.0%) |
| Hilde & Iris | 33 (80.5%) | 7 (17.1%) | 1 (1.9%) | 41 (100.0%) |
| Ricarda & Laura | 22 (81.5%) | 2 (7.4%) | 3 (11.1%) | 27 (100.0%) |
| Sandra & Luzie | 5 (45.4%) | 1 (9.2%) | 5 (45.4%) | 11 (100.0%) |
| Overall | 82 (75.9%) | 11 (10.2%) | 15 (13.9%) | 108 (100.0%) |

TinkerPlots skills, and only approximately 14% of the codings are related to low TinkerPlots skills. Therefore, the participants seem to be capable of using TinkerPlots for their purposes when comparing groups. There are some small differences in between the pairs: Whereas three pairs (Conrad and Maria, Hilde and Iris and Ricarda and Laura) show they have good TinkerPlots skills (percentage of codings related to high TinkerPlots skills are larger than 75%), Sandra and Luzie show some problems in their TinkerPlots use, since only 45.4% of their codings are related to a high TinkerPlots skill and 45.4% of their codings are related to low TinkerPlots skills. These limited TinkerPlots skills also hindered them in their investigation and data exploration process because they were neither able to calculate summary statistics nor to display common graphs for group comparisons (like boxplots) in TinkerPlots.

13.4.7 Relationships Between the Dimensions D1 and D2

To identify any patterns between the pairs in regard to the levels “Statistical reasoning when comparing groups high” and “TinkerPlots skills high”, the percentages of codings of “TinkerPlots skill high” and “Statistical reasoning when comparing groups high” were used (see Fig. 13.12). Laura and Ricarda show good statistical reasoning when comparing groups and also show a competent use of TinkerPlots when comparing groups. The statistical reasoning when comparing groups of Hilde and Iris can be evaluated between high and medium, since 20% of their codings are related to a high and 80% of their codings are related to a medium quality. The software skills of Hilde and Iris can be rated “high”. Conrad and Maria also show high TinkerPlots skills, but they lack statistical reasoning since none of their codings is related to a high quality comparison. So their statistical reasoning when comparing groups would overall be described as medium. Sandra and Luzie show rather medium skills when comparing groups; like Conrad and Maria, they lack statistical reasoning with no codings of high quality. Taking into account the reasoning and the skills of these four pairs, it is not possible to infer from high TinkerPlots skills that there will be high statistical reasoning elements. On the other hand those pairs in our study who show high statistical reasoning elements (like Hilde and Iris and Laura & Ricarda) also tend to show high TinkerPlots skills.

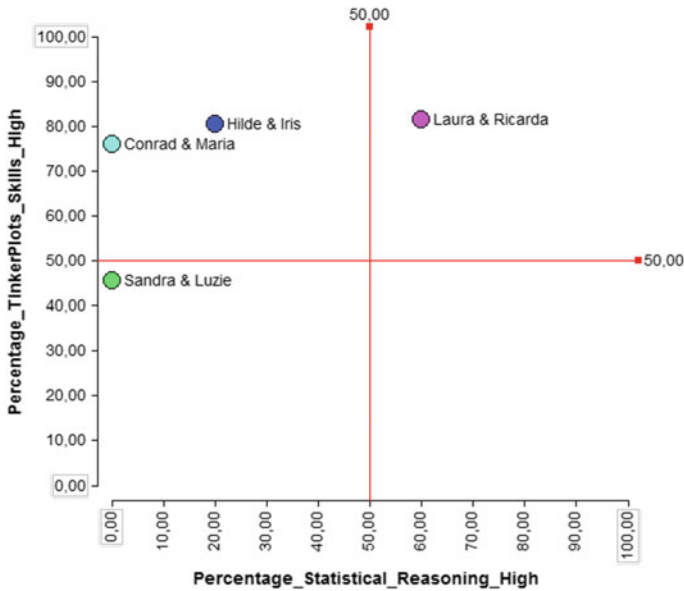


Fig. 13.12 Scatterplot with percentages of codings in regard to “Statistical Reasoning when comparing groups high” and “TinkerPlots skills high”

13.5 Summary and Conclusion

The application of the frameworks D1 and D2 shows that the whole range of comparison elements was used by the participants in the study. The participants showed a competent use of the tool TinkerPlots, because they were able to use the tool for their purposes when comparing groups, were able to explore data with TinkerPlots, and were able to carve out differences between distributions in these large datasets. The comparison statements made were at least on a medium level. On the other hand, the application of the frameworks also shows some discrepancies between the pairs in detail, as mentioned in the results section: The majority of the comparisons (71.4%) are rated as medium quality because the differences were just described but not interpreted. Similar findings were also found in the empirical study of Pfannkuch (2007). Also Biehler (1997) and Frischmeier (2014) noted that learners often concentrate on the production of displays and the calculation of summary statistics but do not interpret their findings. Furthermore—with regard to the norms set in the course (“work out as many differences as possible” and “interpret the differences”)—too many comparison statements lacked interpretation, and some of the pairs like Sandra and Luzie did not use the variety of possible comparison elements to identify the differences between the distributions. In addition only two q -based comparisons were used, although using q -based comparisons was taught in the course when introducing boxplot comparisons. The “shift model” was not used at all, although it was

presented as an adequate strategy to identify shifts between two distributions and it happened to be visible in the data.

The frameworks displayed in Tables 13.1 and 13.2 applied to the data of the video study can help to evaluate the performance of preservice teachers when comparing groups with TinkerPlots. The framework on statistical reasoning gives insight on the one hand to see which group comparison elements are taken into account and on the other hand demonstrates whether the differences between the groups are worked out correctly or whether they are described or interpreted.

The framework of the software skills shows how competent learners can use TinkerPlots in a group comparison process. Independently from TinkerPlots, the framework can also be applied with different software/tools like Excel, Fathom, etc. To rate software skills with this framework the communication of the participants when using TinkerPlots for comparing groups should be taken into account because of the special operationalization of TinkerPlots skills in this framework. The TinkerPlots skills are operationalized in a way that a high skill is given when the intention which was articulated can be fulfilled successfully with the software. To be able to rate the skill, you have to know what the learners have articulated before.

In general both frameworks can help teachers to identify learners' problems with several group comparison elements which might not have been used in the comparison process or which have been used incorrectly. Here for example a teacher could identify that q -based comparisons were not often used and that there was no high quality shift comparison and no high quality p -based comparison, and therefore a teacher could pay more attention to these aspects in further courses. The frameworks can also be used as a norm for teaching. Teachers could teach students to work out differences between two or more distributions in regard to group comparison elements such as center, pointing out the different qualities of comparison (high, medium, low) and trying to enhance interpreting differences between groups. Researchers doing a qualitative study might find the framework useful as a method to analyze data on statistical reasoning in regard to comparing groups.

13.6 Further Research

Areas of further research can be divided in three facets. First the framework can be refined (more dimensions/more levels/more elements) and applied to larger data sets. The framework can also be applied in other settings, for example for analyzing empirical qualitative data on comparing groups activities for primary school. Classroom research in regard to enhancing data analysis with TinkerPlots in primary school with 3rd and 4th graders is now ongoing (e.g. Breker 2016). A third aspect of further research is learning from the results of the video-study in regard to course design (Re-Design of the course "Developing statistical thinking and reasoning with TinkerPlots"). In a re-designed course the norm "work out as many differences as possible" would be set up in a more explicit way with a focus on interpreting worked out differences, which may happen as part of in-class-discussion while considering

adequate and non-adequate examples for interpretation. Additionally q -based comparisons, the “shift model” and the interpretation of p -based comparisons would need to be highlighted in a re-designed course.

References

- Arbeitskreis Stochastik der Gesellschaft für Didaktik der Mathematik. (2012). Empfehlungen für die Stochastikausbildung von Lehrkräften an Grundschulen (Recommendations for the statistical education of teachers at primary school), August 23, 2016, Retrieved from http://www.mathematik.uni-dortmund.de/ak-stoch/Empfehlungen_Stochastik_Grundschule.pdf.
- Batanero, C., Burrill, G., & Reading, C. (2011). *Teaching statistics in school mathematics—challenges for teaching and teacher education*. Heidelberg: Springer.
- Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal*, 3(2), 42–63.
- 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 reasoning of students – concepts and results of empirical studies in the context of group comparisons]. In M. Borovcnik (Ed.), *Anregungen zum Stochastikunterricht* (pp. 97–114). Franzbecker: Hildesheim.
- Biehler, R. (2007). Students’ strategies of comparing distributions in an exploratory data analysis context [Electronic Version]. In *CD-ROM Proceedings of 56th Session of the International Statistical Institute*. https://www.stat.auckland.ac.nz/~iase/publications/isi56/IPM37_Biehler.pdf.
- Biehler, R., & Frischemeier, D. (2015). “Verdienen Männer mehr als Frauen?” - Reale Daten im Stochastikunterricht mit der Software TinkerPlots erforschen [“Do men earn more than women?” Exploring real data in statistics classroom with TinkerPlots]. *Stochastik in der Schule*, 35(1), 7–18.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Blum, W., Driike-Noe, C., Hartung, R., & Köller, O. (2006). *Bildungsstandards Mathematik: konkret*. Berlin: Cornelsen Verlag.
- Breker, R. (2016). *Design, Durchführung und Evaluation einer Unterrichtseinheit zur Entwicklung der Kompetenz “Verteilungen zu vergleichen” in einer 4. Klasse unter Verwendung der Software TinkerPlots und neuer Medien [Design, realization and evaluation of a teaching unit to enhance statistical reasoning in regard to comparing groups in grade 4 using TinkerPlots] (Bachelor thesis)*. Universität Paderborn.
- 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). Heidelberg: Springer.
- Cobb, P., Confrey, J., deSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.
- Frischemeier, D. (2014). 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, July, 2014)*, Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute.
- Frischemeier, D. (2017). *Statistisch denken und forschen lernen mit der Software TinkerPlots [Developing statistical reasoning and thinking with TinkerPlots]*. Wiesbaden: Springer Spektrum.

- Frischemeier, D., & Biehler, R. (2014). Design and exploratory evaluation of a learning trajectory leading to do randomization tests facilitated by TinkerPlots. In B. Ubuz, C. Haser & M. A. Mariotti (Eds.), *Proceedings of the eighth congress of the European Society for Research in Mathematics Education* (pp. 799–809).
- Frischemeier, D., & Biehler, R. (2016). Preservice teachers' statistical reasoning when comparing groups facilitated by software. In K. Krainer & N. Vondrova (Eds.), *Proceedings of the ninth congress of the European Society for Research in Mathematics Education* (pp. 643–650). Charles University in Prague, Faculty of Education and ERME.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning. Connecting research and teaching practice*. The Netherlands: Springer.
- Hasemann, K., & Mirwald, E. (2012). Daten, häufigkeit und wahrscheinlichkeit [Data, frequency and probability]. In G. Walther, M. van den Heuvel-Panhuizen, D. Granzer & O. Köller (Eds.), *Bildungsstandards für die Grundschule: Mathematik konkret*. Cornelsen Verlag Scriptor (pp. 141–161). Berlin.
- Konold, C., & Higgins, T. (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., & Miller, C. (2011). *TinkerPlots™ Version 2 [computer software]*. Emeryville, CA: KCP.
- Kuckartz, U. (2012). *Qualitative inhaltsanalyse: Methoden, praxis, computerunterstützung [Qualitative content analysis: Methods, practice, computer assistance]*. Beltz Juventa.
- 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. 359–370). Dordrecht: Springer.
- Makar, K., & Confrey, J. (2002). *Comparing two distributions: Investigating secondary teachers' statistical thinking*. In Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa. International Association for Statistical Education.
- 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). Dordrecht: Kluwer Academic Publishers.
- Maxara, C. (2009). *Stochastische simulation von zufallsexperimenten mit Fathom - Eine theoretische werkzeuganalyse und explorative fallstudie [Stochastic simulation of chance experiments - A theoretical tool analysis and an exploratory case study]*. Hildesheim: Franz Becker.
- Maxara, C. (2014). Konzeptualisierung unterschiedlicher kompetenzen und ihrer wechselwirkungen, wie sie bei der bearbeitung von stochastischen simulationaufgaben mit dem computer auftreten [Conceptualization of different competencies and their interdependencies when doing stochastic simulations with computers]. 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. 321–336). Wiesbaden: Springer Spektrum.
- Mayring, P. (2001). *Combination and integration of qualitative and quantitative analysis*. Paper presented at the Forum Qualitative Sozialforschung/Forum, Qualitative Social Research.
- Mayring, P. (2010). *Qualitative inhaltsanalyse: Grundlagen und techniken [Qualitative content analysis: Groundwork and techniques]*. Wiesbaden: Beltz.
- Mayring, P. (2015). Qualitative content analysis: Theoretical background and procedures. In A. Bikner-Ahsbals, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 365–380). The Netherlands: Springer.
- Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of boxplots. *International Electronic Journal of Mathematics Education*, 2(3), 149–167.
- Pfannkuch, M., Budgett, S., Parsonage, R., & Horring, J. (2004). *Comparison of data plots: Building a pedagogical framework*. Paper presented at the Tenth International Congress on Mathematics Education (ICME-10), Copenhagen, Denmark.

- Roseth, C. J., Garfield, J. B., & Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education*, *16*(1), 1–15.
- Rossmann, A. J., Chance, B. L., & Lock, R. H. (2001). *Workshop statistics: Discovery with data and Fathom*. Emeryville, CA: Key College Publishing.
- Watson, J. M., & Moritz, J. B. (1999). The beginnings of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, *37*, 145–168.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–265.

Part IV
Teachers' Perspectives

Chapter 14

Teachers' Perspectives About Statistical Reasoning: Opportunities and Challenges for Its Development



Helia Oliveira and Ana Henriques

Abstract This study concerns the perspectives of 11 mathematics teachers about the potential and the challenges of developing a learning environment targeting students' statistical reasoning in a Developmental Research Project context. We focus on their perspectives regarding the tasks, the software, and their role in promoting an adequate classroom discourse, from their written answers to an open questionnaire. Findings show that the teachers distinguish key aspects that characterize the statistical reasoning conveyed by the tasks and ways the students used the software to explore them, as well as the necessity of assuming a new role that stands in contrast with traditional teaching practices. The findings also point out several obstacles that give rise to a reflection about the design of the project.

Keywords SRLE · Statistical reasoning · Tasks · Teachers' perspectives
Technology

14.1 Introduction

International current guidelines for teaching statistics (e.g. Franklin et al. 2007) prioritize the development of students' literacy and reasoning and challenge teachers to create learning environments that are in contrast to the prevailing classroom practices. Given that the classroom environment and the adopted teaching approaches, including the tasks to be proposed, are key factors in learning, Garfield and Ben-Zvi (2009) offer a model for a Statistical Reasoning Learning Environment (SRLE) to support a thorough understanding of statistics and the development of statistical reasoning. However, the adoption of new curricular perspectives and the need to promote learning environments with such characteristics is a novelty to teachers whose

H. Oliveira (✉) · A. Henriques
Instituto de Educação, Universidade de Lisboa, Lisbon, Portugal
e-mail: hmoliveira@ie.ulisboa.pt

A. Henriques
e-mail: achenriques@ie.ulisboa.pt

experience is mainly based on descriptive statistics (Ben-Zvi and Makar 2016) and raises a number of challenges for them (Leavy 2010; Makar and Fielding-Wells 2011). Thus, teacher educators are required to create opportunities for teachers' professional development in statistics, particularly through courses that are closely related to the teaching practice (Ben-Zvi and Makar 2016). Several studies show that, when implementing new approaches in their lessons, teachers may benefit from collaborative working environments involving other teachers and researchers (Goodchild 2014; Makar and Fielding-Wells 2011; Potari et al. 2010).

To promote teaching practice under the SRLE model with a group of Portuguese mathematics teachers, a Developmental Research Project (DRP) was conducted where teachers and researchers worked collaboratively in the design and implementation of tasks that seek to promote students' statistical reasoning. Recognizing that teachers do not have many opportunities for professional development in statistics (Batanero et al. 2011) and that the envisioned teaching practice is rather different from the traditional one (Garfield and Ben-Zvi 2009), a study to understand the teachers' perspectives about their lived experience with SRLE in the project was carried out. Acknowledging that research on teachers' education for these new approaches in statistics is still scarce (Ben-Zvi and Makar 2016), and that the important dimensions of teachers' beliefs and attitudes have received less attention from the research than teachers' knowledge (Batanero et al. 2011; März and Kelchtermans 2013), this study aims to contribute to the knowledge about how innovative teaching practices targeting students' statistical reasoning are perceived by the teachers.

Recognizing the novelty and complexity of the envisioned practice according with the SRLE model (Garfield and Ben-Zvi 2009), it is important to understand how teachers perceived the environment's specific features. We focus our study on the teachers' perspectives concerning the tasks and the software that was used and their own role as teachers in promoting an adequate classroom discourse, which are three of the main elements of SRLE that have received particular attention in the DRP. In accordance with our research goal, we formulated the following research question: What are the teachers' perspectives about the potential and the challenges of developing a learning environment targeting students' statistical reasoning?

14.2 Theoretical Framework

14.2.1 *Statistical Reasoning*

Statistical reasoning, commonly described as the way individuals reason using statistical ideas and how they give meaning to statistical information, encompasses the conceptual understanding of important ideas such as variation, distribution, center, spread, association and sampling or the combination of ideas about data and uncertainty that leads to inference (Ben-Zvi and Garfield 2004; Makar et al. 2011). This ability to draw conclusions that extend beyond the available data has received

increasing attention in both curriculum documents and research in statistics education, and Informal Statistical Inference (ISI) has become a key objective of statistical reasoning.

Students' informal ideas of statistical inference should be developed from the first years of school as it is a known area of difficulty for older students when formal ideas are later introduced (Franklin et al. 2007; Makar et al. 2011; Watson 2008). However, ISI should not be taught to students as an entity in itself, but rather it would be preferable to focus the instruction on reasoning processes that lead to inference (Makar and Rubin 2009). Such informal inferential reasoning means "drawing conclusions or making predictions, informally, on a 'broadened universe' from patterns, representations, statistical measures and statistical models of random samples, while considering the strength and limitations of the sample and the inferences reached" (Ben-Zvi 2006, p. 2).

Recent research focused on ISI supports the use of statistical investigations in classrooms in order to foster the emergence of students' inferential practices (Henriques and Oliveira 2016; Leavy 2010). Thus, students can experience statistics as an investigative process to solve real problems, as they are encouraged to formulate their own statistical questions (conjectures) about a significant phenomenon, to plan a proper data collection, to select suitable graphic and numerical methods in order to analyse such data and to draw conclusions and inferences from the developed activity (Franklin et al. 2007). Due to their nature, statistical investigations often provide a distinctive context for observing students' conceptual ideas about statistical reasoning, namely fundamental processes like variation, transnumeration, evaluating statistical models and integrating contextual and statistical features of the problem (Wild and Pfannkuch 1999). At the same time, they may involve students in fundamental components of informal inference, such as decision making and prediction (Makar and Rubin 2009; Watson 2008). For teachers, the use of investigations also provides knowledge that can be used in the design, implementation, and assessment of instruction in statistics and data exploration (Henriques and Oliveira 2013), since they incorporate domain-specific knowledge of students' statistical reasoning.

The huge development of technology and the accessibility of real data have an important impact on the curriculum guidelines, providing students and teachers with new tools to explore the ISI in rich and meaningful contexts, including the use of a broader process of statistical investigation (Ben-Zvi et al. 2012). In particular, the use of dynamic statistical learning environments, such as *TinkerPlots*TM (Konold and Miller 2005), in combination with appropriate curricula and instructional settings, has shown great potential in the learning of statistics and in the development of students' statistical reasoning, particularly the Informal Inferential Reasoning (Makar et al. 2011). This chapter highlights the possibility of using databases already integrated in the software reducing the amount of time to collect data, as an integrated stage of the investigation cycle. In addition, students can create and explore the potential of their own graphical representations and statistical measures taking advantage of the software tools. This helps students to actively build knowledge "doing" and "seeing" statistics and to think about observed phenomena (Konold and Miller 2005). The dynamic nature of the software also enables young students to informally explore

data to make conjectures and to use the experimental results to test or modify these conjectures (Paparistodemou and Meletiou-Mavrotheris 2008). Discussion is stimulated, as the results of prediction or conjectures are rapidly viewed, and allows students to look for justifications for their statements, reinforcing their ability to reason statistically (Ben-Zvi 2006; Watson 2008).

In line with these curricular orientations, the Portuguese mathematics syllabus for basic education (ME 2007) advocated a more in-depth and extended role of statistics in school mathematics suggesting the use of data-oriented approaches and statistical investigations in the teaching and learning of this topic for all grades from primary to lower secondary school (6–14 years old). However, since in the national context formal statistical inference is reserved for university courses and, traditionally, students are not exposed to ISI methods before that, informal inferential reasoning is not explicitly referred as a learning objective in that curricular document. Nevertheless, the same document suggests that activities should be conducted to promote the emergence of this reasoning by provoking students to “ask questions about a certain subject, identify [relevant] data to be collected, and organize, represent and interpret such data in order to answer the questions previously asked” (p. 26). In these situations, the teacher has the responsibility to encourage students to make decisions and inferences based on the collected data and to use proper language, considering their development level (ME 2007). This document also recognizes the important role of technology in data handling, justifying that technological tools “are fundamental (...) to carry out work under this topic, as they allow students to focus on choosing and supporting the methods to be used in data analysis and results interpretation, releasing them from time consuming calculations” (p. 43). Despite these recommendations, statistics has received little attention from mathematics teachers in basic education in Portugal. In addition, the integration of technology in education is not yet widespread, due to the lack of training in this area and to countless difficulties teachers face because of the scarcity or absence of technology in a large number of schools.

These new perspectives on teaching and learning of statistics call for changes in the teaching practices, namely in regard to the contexts to be presented to students, the statistical processes to be explored, and the technological resources to be used. In this respect, Garfield and Ben-Zvi (2009) point out a statistical learning environment that promotes the development of statistical reasoning (SRLE), based on six principles of instructional design:

- (1) focus on developing central statistical ideas, such as distribution, variability, covariation, sampling and inference, in order to deepen the conceptual understanding of students instead of learning procedures;
- (2) use of real data, challenging students to become involved in data collection and in the formulation of conjectures and statistical inferences based on data analysis;
- (3) use of classroom activities to develop students’ statistical reasoning, especially focusing on the proposed tasks and on how to develop them;

- (4) integrate appropriate technological tools to help data exploration and analysis, supporting students in results' interpretation and conceptual understanding;
- (5) promote classroom discourse focused on meaningful statistical ideas, promoting discussion and negotiation of meanings; and
- (6) use of assessment to monitor students' learning, focusing on understanding rather than on skills, and to reflect about the learning process.

14.2.2 Teachers' Perspectives

Teachers' practices are influenced by many factors. Teachers' perspectives on statistics as an important factor influencing their willingness to adopt new teaching strategies and, therefore, the successful implementation of new curriculum guidelines, has long been established (Estrada et al. 2011). Recent research on non-cognitive factors such as teachers' beliefs and attitudes towards statistics and its teaching gives some sense of teachers' perceptions about factors that determine or affect their teaching practice. For example, in Martins et al. (2012) study, basic education school teachers enjoyed teaching statistics and agreed to include the topic in the curriculum since they recognized its importance and usefulness in students' daily life. Findings from previous studies (Chick and Pierce 2011; Watson 2001) also showed that teachers, in general, have positive attitudes towards statistics. However, in Watson's (2001) study, most of the teachers admitted the need for professional development in order to improve their ability to teach statistics in accordance with the new perspectives, favouring classroom-based work and collaborative contexts.

To study teachers' perspectives on statistics, it is fundamental to know their perceptions regarding new recommendations for teaching statistics, such as GAISE (Franklin et al. 2007). These perceptions have become a focus of analysis, and Groth (2008) noted that teachers seldom see these reforms as a means to significantly change the teaching objectives and their approaches to this subject. Instead, they tend to perceive these reforms as supplements or revisions of their current pedagogical repertoire. For instance, in the studies conducted by Groth (2008) and Chick and Pierce (2011), teachers from different education levels, who participated in a focus group about the implementation of GAISE recommendations, seem not to be aware of the need to develop their knowledge about how different statistics is from mathematics. Regarding the use of technology, Groth (2008) also concluded that the teachers are aware of such recommendations and that their positive attitudes contribute to its successful integration in the teaching and learning processes. However, according to the author, teachers might use technology superficially only to follow the syllabus, if they do not understand how to use it properly to teach a specific statistical topic. The same difficulties were present for other educational strategies such as the use of cooperative learning. Therefore, assessing teachers' perceptions regarding the proposed GAISE reforms may provide opportunities to disclose both the difficulties that emerge while those recommendations are implemented and some lines of action for teacher education.

14.3 The Developmental Research Project

14.3.1 Context and Participants

This study falls under the scope of a Developmental Research Project (Goodchild 2014) aimed to build and test a sequence of tasks oriented towards students' statistical reasoning, using *TinkerPlots*TM software. Based on a Design Research perspective (Cobb et al. 2009), involving interactive cycles of planning, implementation and reflection, the project intended to build professional development opportunities in statistics for Portuguese mathematics teachers of basic education (5th to 9th grades). The 11 teachers who participated in the DRP taught in public schools located within urban or suburban areas mainly with disadvantaged socioeconomic conditions, where many students come from single-parent families and have a school retention background. These teachers were known by the two researchers (the authors) as effective professionals who had some affinity with an inquiry approach in mathematics teaching (Menezes et al. 2015) and became interested in learning how to develop an instructional approach aligned with the recent curricular trends in school statistics (Franklin et al. 2007; ME 2007). The topic of statistical reasoning, particularly the Informal Inferential Reasoning (IIR), and the use of *TinkerPlots*TM were a novelty for the majority of them. Considering the importance of the classroom environment and the adopted teaching approach in association with the challenging tasks proposed to students, the researchers first discussed the SRLE perspective with this group of teachers (Garfield and Ben-Zvi 2009). In the subsequent meetings, the participant teachers and the authors were involved in the design of tasks and resources for a sequence of lessons in line with the SRLE principles, and using *TinkerPlots*TM.

The DRP was developed during one school year, between November and June, with 40 h of meetings of joint work and assuming an essentially collaborative nature. The authors undertook the double role of researchers and teacher educators, and the teachers were co-responsible for the proposition and discussion of tasks, classroom implementation and reflection about the whole process.

The mathematics teachers worked in pairs or in small groups of three in the planning of the lessons, collecting data and reflecting on them. One or both teachers in the pair implemented the sequence of tasks with their own class, with the support of the other teacher for the classroom management and data collection. After finishing the tasks' sequence with their students, the teachers shared and discussed their experience in the DRP meetings. The materials produced as well as the joint reflection in the group informed the tasks' reformulation and the conditions for their implementation in the subsequent cycle. In the last two meetings of the DRP, each group of teachers presented orally a report about the evidences of their students' IIR, discussed it with the larger group, and afterwards produced a short written report concerning the same issues.

14.3.2 *Teaching and Learning Environment: Tasks and Software*

A sequence of three tasks solved in 8th grade classes and the way they were explored by the students is presented to illustrate the worked carried out in the DRP. The tasks were designed to be solved in two or three lessons of 90 min. Although these tasks were not intended to cover all the statistical topics included in the syllabus for the considered grade, they were aligned with the Portuguese mathematical syllabus (ME 2007). As statistical inference is not a learning objective in the mathematics syllabus, the tasks were not explicitly designed to focus on inference. However, they included a set of questions that embody the IIR components described in literature (Makar and Rubin 2009; Zieffler et al. 2008) and therefore provided opportunities to engage and simultaneously support students in several aspects of informal inferential practice (Henriques and Oliveira 2016). For example, when solving these tasks, students could use their intuitive or prior knowledge on fundamental statistical concepts and probabilistic language to make predictions, without using formal statistical methodology, and were challenged to give justifications based on evidence. *TinkerPlots*TM software was used in all tasks, as a tool for data handling, since it is easy to use and provides a dynamic learning environment to support the development of students' statistical reasoning.

Tasks 1 and 2 were applied to build the skills students would need in Task 3 for conducting a statistical investigation and for making inferences, as they were not familiar with this kind of activity nor had previous experiences with technology in their mathematics lessons. Task 1, *Nenana* (Fig. 14.1), is the one with a more structured nature and starts from a research question, being guided by a set of questions that lead students to interpret a real context situation by exploring the available data in *TinkerPlots*TM. The students can use different representations when exploring these data in order to obtain evidence for their claims and to make predictions in several ways. The task also leads students to discuss the unpredictability of the phenomena and to become familiar with the notions of distribution and variability, which can also help them to have a critical attitude when confronted with numerical information.

The second task of the sequence, *Task 2, Fish Experiment* (inspired by Bakker and Derry 2011, Fig. 14.2) is aimed to encourage students to think about the research process in statistics and to make inferences from different samples in order to answer an initial question. The first question raised a discussion in class about how to plan a statistical experience which includes obtaining the necessary evidence to verify, based on the data, the truth of the statement in question. To help students understand the need and the advantage of using a representative sample of the population and other factors that determine the precision of any inferences, aspects related to sampling were discussed, for example, which data were needed and how to collect them. Another aspect covered by this task was exploring the notion of sample variability. Using the simulation functionality available in the software, the students were able to create several samples (of growing dimension or with the same dimension) and compare them using, for example, the boxplot tool.

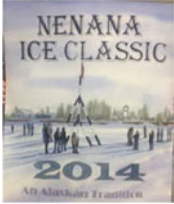
| | |
|---|--|
| TASK 1 – Nenana Ice Classic |  |
| The Nenana Ice Classic is a competition held annually by the town of Nenana, Alaska, to guess when the ice of the Tanana River breaks. People bet on the exact minute that a wooden high tripod will fall into the icy river. The contest was initiated by the inspectors of railways in Alaska in 1917 and has taken place every year since then. In 2015, for example, the winner received a prize of three hundred and thirty thousand US dollars. | |
| In 2001, two Stanford scientists published a paper in the journal <i>Science</i> where they concluded that at that time the thaw occurred 5.5 days earlier than in 1917, when the contest began. Does the thaw occur earlier over the years? | |
| What do you think is the answer to this question? Explain what may support your answer. | |
| We propose you to explore this issue from a database of <i>TinkerPlots</i> software where you can access information about the thaw in the river. Data are available for some attributes (variables) that you can view in the form of cards or tables and whose description is presented there in a text box. | |
| Part I | |
| 1. Check your initial conjecture using a graphical representation of the data on the number of days occurring since January 1 until the day of thawing, over this period of years. What do you think is the answer to this question, now? Justify it. | |
| Part II | |
| You also find information in <i>TinkerPlots</i> on the time of day on which the thaw occurred ("segment" attribute) over the years. Answer the following questions based on graphical representations and justify them. | |
| 1a) In how many years has the thaw occurred in the morning? | |
| b) At what hour of the day does the thaw occur more often? Explain your thought. | |
| 2a) What can you conclude about the different months in relation to the time of day when the thaw occurs? What evidence do you have about that? | |
| b) And what can you conclude about the different months in relation to the month in which the thaw occurs over the years? What evidence do you have about that? | |
| 3. Consider the following statement: "The thaw in May tends to occur earlier during the day than in April". | |
| a) Use a diagram of extreme and quartiles (based on Boxplot tool <i>TinkerPlots</i>) to check the veracity of the above. Draw the diagram on your sheet. | |
| b) Based on the information that can you take from the diagram of extreme and quartiles and explain why you agree or disagree with the above statement. | |


Fig. 14.1 Task 1, *Nenana*

Finally, Task 3, *The human body: a study in school* (Fig. 14.3), engaged students in all phases of a statistical investigation (Wild and Pfannkuch 1999) to discover more about the students in their school. Students were required to use their previous knowledge about the context and statistical concepts and processes (e.g. understanding the need for data, graphical representations, distribution, variability, covariation, and sampling) as well as probabilistic language to make informal judgments and predictions about the school population, based on data collected in their class. The task also challenged students to explain their reasoning, integrating persuasive data-based arguments in their conclusions. This work was supported by the exploration of their real data through various representations, facilitated by the use of *TinkerPlots*TM.

In line with the two first principles of SRLE (Garfield and Ben-Zvi 2009)—developing central statistical ideas and using real data—the three tasks aimed to develop diverse central statistical ideas and intended to get students involved in the exploration and analysis of a set of real data. The data sets were provided directly or obtained by students through simulation or data collection, aiming towards understanding the need for data in order to draw conclusions. The visual analysis of these data, encouraged by many of the formulated questions in the tasks, could lead students to develop the notions of covariation (informally) and distribution, to recognize data variability and to draw together dispersion measures with central tendency in

TASK 2 – Fish Experiment

Aquaculture is the farming of aquatic organisms such as fish, molluscs, crustaceans and amphibians as well as the cultivation of aquatic plants for human use. This activity is operating for a long time, as there are records referring to aquaculture practices by Chinese people several centuries before our era and to the growth of Nile tilapia (*Sarotherodon niloticus*) 4000 years ago. Currently, aquaculture is responsible for producing half of the fish consumed by the world population. (<http://en.wikipedia.org/wiki/Aquaculture>)



A fish farmer has stocked a pond with a new type of genetically engineered fish. The company that supplied them claims that “genetically engineered fish will grow to be longer, reaching twice the length of normal fish”.

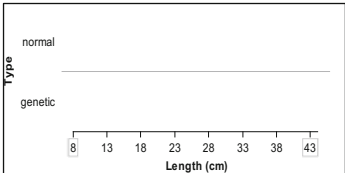
Do you think the fish farmer can rely on the claim of the company? What should he do to check its veracity?

You now have the opportunity to simulate a ‘catching’ to answer the question: *Do genetically engineered fish grow longer than normal fish? If so, how much?*

Part I

The fish farmer decided to stock their pond with 625 fish, some tagged as normal and some as genetically engineered. After they were fully grown, the farmer caught fishes from the pond and measured them. In your experiment, those fish are simulated by data cards (fish-card) which show their type (normal or genetically engineered) and length (in cm).

- In the graph below, record the data (fish type and length) of a set of 25 fish-cards resulting from the simulated ‘catching’ carried out by you and your colleagues.
- Based on that graphical representation, try to answer the question: *Do genetically engineered fish grow longer than normal fish? If so, how much?*
- Would your answer be the same if you had selected one sample of size 50? Explain why.



Part II

We now propose an exploration on this question from a database of *TinkerPlots* software – “Fish Experiment”, where you can simulate samples of different sizes (SampleSize = 25, 30, 50, 75, ...).

- Simulate three samples of sizes 25, 50 and 100. For each of them, based on the resulting graphical representation, create a boxplot and record, in the table below, some of the statistical measures you get.

| | | Growing samples | | |
|------------------------------------|----------|-----------------|---------------|----------------|
| | | Sample 1 (25) | Sample 2 (50) | Sample 3 (100) |
| Genetically engineered fish | Mean | | | |
| | Median | | | |
| | IQ range | | | |
| Normal fish | Mean | | | |
| | Median | | | |
| | IQ range | | | |

- Based on the recorded information, compare the distributions of the diverse samples obtained and answer to the initial question of the task. Give arguments to justify your response and to help the fish farmer to decide whether to keep doing business with the company.

Part III

We propose another challenge: *Will several samples of the same size be similar to each other?*
 Start your work by simulating a first sample of a size of your choice and then, with the function DATA-> Rerandomize, generate two other samples of the same size. Compare the three samples, using graphical representations and statistical measures that you can obtain and record them in the following table.

| | | Sample size _____ | | |
|------------------------------------|----------|-------------------|----------|----------|
| | | Sample 1 | Sample 2 | Sample 3 |
| Genetically engineered fish | Mean | | | |
| | Median | | | |
| | IQ range | | | |
| Normal fish | Mean | | | |
| | Median | | | |
| | IQ range | | | |

Write a short essay to present your findings regarding the above question and to explain what evidence you have to answer the way you did.


Fig. 14.2 Task 2, *Fish Experiment*

TASK 3 – The human body: a study in school

The *Vitruvian Man* is a famous drawing by Leonardo da Vinci, around 1490, included in his diary. The drawing depicts a man in two superimposed positions with his arms and legs apart and inscribed in a circle and square.

It is based on a famous excerpt of the ancient Roman architect Vitruvius' *Book III* of his treatise *De Architectura*, describing the ideal human proportions. For example, that book specifies that:

- a palm is four fingers
- a foot is four palms
- the length of the outspread arms (arm span) is equal to the height of a man



Vitruvius had already tried to fit the human body proportions in a circle and a square but his attempts were not perfect. It was Leonardo da Vinci who correctly adjusted it within the expected mathematics patterns. ([http://pt.wikipedia.org/wiki/Homem_Vitruviano_\(desenho_de_Leonardo_da_Vinci\)](http://pt.wikipedia.org/wiki/Homem_Vitruviano_(desenho_de_Leonardo_da_Vinci)))

How could you characterize the middle school students in your school regarding some of Vitruvius' measures such as height, foot size and arm span?

Part I

1. Think about what information will be needed to answer this question and how to collect data by answering the following questions.
 - a) What is the population under study?
 - b) What sample size can we work with?
 - c) How could you choose a representative sample?
 - d) What variables should we study? Are these variables qualitative or quantitative? Are they continuous or discrete?
2. Indicate a procedure that would lead to the selection of a biased sample (unrepresentative).
3. For convenience, today we are only going to collect data on our class. Carefully measure and record each attribute (height, arm span, foot size) for each class member.

Part II

1. Analyse the data you collected, which is already in a *TinkerPlots* database.
 - a) What interesting questions about this information could you ask? Consider the following:
 - Students' height;
 - Boys and girls' arm span;
 - Relationship between students' foot size and height;
 - Another aspect that you think is relevant to study.
 - b) What do you think is the answer to your questions? Explain the reasons for your answer.
 - c) Respond to two of the questions asked in 1a) using graphical representations.
2. From the data collected on your class, prepare a short essay on what you could say about the characteristics of all the students in the school, considering the aspects in question 1.a). Explain what evidence you used to make your predictions.

Fig. 14.3 Task 3, *The human body: a study in school*

order to informally describe and explore the presence of variability that arises from sampling and its relevance in inference making. The following three principles of SRLE, developing statistical reasoning, meaningful discourse and using appropriate technology, are related to the classroom activity. The tasks had a central role in the work developed by students in small groups, with a special focus on the collective discussion and the teacher's systematization (Menezes et al. 2015). The use of *TinkerPlots*TM software was a characteristic of the learning environment. The simplicity of creating diverse graphic representations and the instant calculation of statistical measures allowed students to focus on the reasoning processes. Besides, considering the importance of communication in the activity with the software, the students were organized in small groups in front of the computer. In general, teachers also promoted a discourse focused on meaningful statistical ideas that encouraged the argumentation and negotiation of meanings through questions formulated during the tasks, requesting students to show arguments in order to support their inferences. Lastly, regarding the sixth principle, related to using assessment, and considering this as a

research and development project, several tools were used to analyze diverse aspects of the statistical reasoning of the students. The participating teachers collaborated in a joint and continuous reflection about the suitability of the tasks and the way they were implemented (Oliveira and Henriques 2015).

14.4 Methodology

This study follows a qualitative approach (Denzin and Lincoln 2005) focusing on the perspectives of the 11 participating teachers about their experience in the DRP, regarding the promotion of a learning environment that targets students' statistical reasoning. The study reported in this chapter relies specifically on data collected from nine open questions of a questionnaire (Fig. 14.4), individually answered by the teachers, at the end of the project. Since teachers could answer the questions at home and use as much time and space to write as they wanted, the assumption is that they could express their views without major constraints. In order to prevent conditioning the participants' answers, the principles of SRLE (Garfield and Ben-Zvi 2009) were not explicitly referred to in the questionnaire.

Teachers' answers were analyzed regarding their perspectives on: (i) the tasks; (ii) the technology used in these lessons; and (iii) the teacher's role in promoting an adequate classroom discourse. For each of these dimensions, the potential and the challenges that the teachers (whose names are fictitious) mentioned in connection with the SRLE principles (described in Sect. 14.2.1) are given. A first version of the findings was produced which included all quotes that have been identified, followed by a new simplified version keeping the quotes that seemed more clear in the message and eliminated all the others. Finally, the data were checked for consistency between the teachers' answers in the questionnaire (Fig. 14.4) and their group's written reports. No major inconsistencies were found, although the questionnaires provided more information regarding the teachers' perspectives than the reports, as the reports had a more limited focus.

| The questionnaire | |
|--------------------------|--|
| 1. | In your opinion, what is importance of developing students' statistical reasoning? |
| 2. | Regarding the class where this experiment was carried out, please indicate what aspects of statistical reasoning were shown by the students and the ones that have proved to be more difficult for them. |
| 3. | In your opinion, what is the role of the teacher when promoting the student' statistical reasoning? |
| 4. | What difficulties have you faced as a teacher to promoting students' statistical reasoning in the lessons where these tasks were applied? |
| 5. | In your opinion, what are the main characteristics of a task aiming to develop students' statistical reasoning? |
| 6. | What aspects stand out regarding the use of TinkerPlots in solving these tasks in the classroom? |
| 7. | Are you planning to use TinkerPlots again in the near future? Why? |
| 8. | To what extent do you feel to be prepared to continue developing this kind of work with your students? Explain? |
| 9. | Did you develop new perspectives on the teaching and learning of Statistics during this experience? Explain. |

Fig. 14.4 The teachers' questionnaire

14.5 Findings

14.5.1 *The Tasks in the Project*

The teachers' answers reflect their positive assessment of the tasks developed under the project and implemented in their classrooms. All teachers recognized the utility of statistics for their students, and most of them seemed to be aware of the specificity of the subject regarding the centrality of data and the key ideas and statistical processes involved. For example, Alice, highlighted the importance of the data in statistics associated with a context, by stating that "In statistics data are regarded as numbers in context, which, in turn, gives meaning to the interpretation of the results but also constrains the procedures", a noticeable aspect during the tasks' exploration by the students. Some teachers also underlined that working with these tasks in their lessons conveyed to them a new vision about the distinguishing characteristics of statistics and in particular of statistical reasoning: "Realizing that statistical reasoning is not the same as mathematical reasoning, and that they are promoted in different ways, has been a new perspective presented to me by this project" (Clara).

The participant teachers proved to be very responsive to the use of real data in the proposed tasks, which they considered of paramount importance to involve students in the work and to support their reasoning. This is referred to by all the teachers, indicated by Andreia's words: "The exploratory tasks should be based on real contexts in order to involve students but also to support the development of their inferential reasoning". Some teachers also pointed out that it would be beneficial if the data were collected by the students themselves considering that, with such activity, "their commitment" (Rosário) could be improved.

These teachers also recognized the proposed tasks' potential for developing statistical ideas and pointed out, with different emphasis, the need for data and other ideas such as distribution, data noise and center, covariation, sampling and inference. Carlota, for example, referred to the tasks' characteristics, which were aimed to promote students' statistical reasoning: "They should first help students to make conjectures so that, with the sample, the students can test such conjectures and organize the data in order to support their claims". On the other hand, another teacher, Catarina, mentioned the diverse statistical processes in which students were involved. For example, in Task 1 (Fig. 14.1), presenting data in different ways, reasoning with statistical models and integrating the context in the statistical analysis and, in Task 3 (Fig. 14.3), acknowledging the need for data and recognizing data noise:

I think students showed, during the work developed in task 1 (...), an evolution in the following components – transnumeration, reasoning with statistical models and integration of the context in the statistical analysis. (...) Lastly, in the third task (...), I think it is possible to say that the mostly highlighted SR components were: the need for data, recognizing variability and reasoning with statistical models.

As mentioned above, the proposed tasks led to the exploration of data in ways that students could develop the informal notion of covariation. Some teachers referred to

the informal representations used by the students to explore the existence of relationships between variables [in task 3, Fig. 14.3], made possible by the software, as one of the aspects covered in these lessons. For instance, Maira stated: “[the software] allows [exploring] the possible connection between two variables. It allows students to choose their own informal representations”.

Distribution, another key notion aimed to be developed with these tasks by the creation and exploration of graphics in the *TinkerPlots*TM, is also implicitly referred by the teachers. The idea of distribution center, summarized in a statistical measure, is cited by several teachers as an evidence of students' meaningful activity and learning experiences in statistics; Isabel stated, “A large number of groups used mean and median in order to generalize the information, with others considering the interquartile range a good indicator”. However, these teachers did not mention any joint interpretation of a central tendency measure and a spread measure. Sampling and informal inference were the most targeted ideas in the tasks in the project. The former, because it is part of the mathematics basic education syllabus in Portugal, and the latter because it is considered a central aspect of the statistical reasoning. Some teachers' opinion is that though these are complex notions for the students they were able to grasp them, at least to a limited extent. In fact, as stated by Susana, in many cases “Students showed they can make generalizations beyond the data as well as use data as evidence to make generalizations”.

The understanding of the sampling process is scarcely stressed by the teachers, but still some of them referred it when claiming that the work, in particular with the software, helped students to understand important concepts such “as sample, data noise in a sample process and variability in equal size samples from a certain population [in Task 2, Fig. 14.2]” (Catarina).

Some of the main challenges pointed out by the teachers regarding the statistical reasoning targeted by the tasks are also connected with the same big statistical ideas mentioned above. For instance, some teachers argued that students tended to disregard data noise in the tendencies that were to be generalized and that, in general, they did not use the probabilistic language or the uncertainty concept in generalizing processes, as referred by Rosário:

Most of the time, the biggest difficulty [for students] is to move from data to bigger sets using the probabilistic language. This transition and its discovery were always difficult to be achieved [by students] mostly due to the difficulty to understand the concept of uncertainty, which is typical of statistics and probabilities (a topic that will be taught in the 9th grade). For these [8th grade] students, mathematics was still perceived as an “exact truth” producer and the introduction of the uncertainty notion was a novelty.

This last difficulty may be associated with the little attention given to it when planning the sequence of lessons in the DRP and supporting students' the work in the classroom, namely during the tasks final discussion involving the whole class. The established goals for these lessons, centered on the promotion of students' statistical reasoning, defied the teachers' practice, as Clara noted: “The development and approach of tasks promoting inferential reasoning is a great challenge, and it requires some reflection and consolidation of the fundamental ideas [on part of the teacher]”. According to this teacher, for continuing working with the students with

respect to these ideas “a further study on the meaning of statistical reasoning and its development with the students” would be necessary.

While the teachers valued the presence of real data in the tasks, as mentioned above, they also considered that the exploration of real situations or situations close to reality raised some problems for some students’ reasoning process. This happens when students use their knowledge about the situation to draw conclusions not supported by data, as referred by one teacher regarding the first task of the sequence: “Students made (...) generalizations beyond data (...) when they argued based on their daily knowledge, and not based on data, about [what they expect to happen] in the next 10 years [bringing to discussion] (...) the environment issue [and] global warming” (Penélope). Thus, a central choice option for these tasks, one the teachers consider of high value, may also represent, from their point of view, an obstacle to student’s reasoning and therefore something that they must learn how to deal with in their teaching practice.

14.5.2 *TinkerPlots’ Use by Students*

The software used in all lessons for solving the tasks was acknowledged by all teachers as extremely valuable for creating a learning environment that fosters students’ statistical reasoning. One of the software’s main potentials mentioned by the teachers is the easiness of producing a great variety of graphical representations, almost instantaneously, which allows students to analyze data in many ways and to further develop their understanding of distribution, data noise and covariation, albeit informally as it is intended at this school level. As stated by Isabel, students may therefore “analyze the main characteristics of a distribution (...) and to compare more than one set of data”.

Additionally, due to the fact that students do not waste time in calculations of statistics measures nor in constructing graphical representations by hand, the teachers argued that they could concentrate their efforts “in [data] analysis and in drawing conclusions, [namely] producing informal inferences” (Catarina). They still stressed that this resource favors the formulation and test of conjectures and the discussion and argumentation of ideas, as stated by Andreia: “[*TinkerPlots*TM] makes the lessons more dynamic, providing the opportunity for a more in-depth discussion where students can test and, thus, find evidences to validate their conjectures”. Another highlighted potential of the software by the teachers is that of producing virtual simulations (in Task 2, Fig. 14.2) and to use data from as many samples as one wishes and thus “contributing to students’ reasoning about sample distribution” (Isabel).

The majority of these teachers considered that students started to use *TinkerPlots*TM with relative ease. However, they also mentioned a number of challenges related to its use in the classroom, in particular due to the large number of students in the classes. Some of them stated they had to meet many requests from students regarding the functionalities of the software or the interpretation of the obtained representations and their relationship with the task’s questions, as referred by Susana: “It

was difficult to answer all the students' questions in time, which were often related firstly to how the software worked and then with the diversity of representations and meanings". The challenges connected with the use of the software that was intended to support students' statistical reasoning can also be understood from the teachers' recognition of the need to deepen their knowledge about its use in order to be able to support the students more effectively, as stated by Alice:

At the moment, I feel I have basic training, which allows me to keep working with the students. I need to deepen my knowledge about the software in order to answer more easily the multiple requests made almost simultaneously by the students during the lessons.

We note that teachers are aware of how this technology may help students to explore and interpret the data and to support reasoning processes in statistics. Simultaneously, they also recognize the need to further develop their skills in using *TinkerPlots*TM in the classroom, which is expected given the reduced number of lessons that were conducted.

14.5.3 The Teacher's Role

Besides the tasks and the software used in the DRP, the teachers also recognize the importance of their own role in creating an environment that supports students' statistical reasoning. Susana, one of the teachers who pointed out that different moments of the lesson (task introduction, autonomous group work, and whole class discussion of the students' work) demand different roles, stated that the teacher supports the students "by asking guiding questions, encouraging data analysis, data exploration and the understanding of statistical concepts/representations as well as the written record of conjectures and conclusions, together with justifications". This kind of classroom discourse is referred to by almost all teachers, who recognize its importance in promoting students' statistical reasoning. Carlota argues that: "helping the groups [of students] to argue and justify their answers and the various explorations they did until reaching an answer is fundamental to help other students to understand the "why" concerning the conclusions that are being presented". Participating in this project stimulated the teachers to be more conscious that learning could be promoted by a classroom dynamics focused on the students' activity and in its discussion, as it is the case of Catarina: "I believe I grew greatly with this experience and became more aware on how I can use certain aspects or students' answers to promote their learning from the tasks I will propose".

Teachers also underlined several challenges they faced when conducting these lessons, particularly following up the autonomous work of the groups. In some cases, such challenges were associated with the intention of providing non-directive support in classes with a large number of students. As Carlota mentions, the assistance she would like to provide each group would take too much time, considering the needs of the whole class:

It is not easy to help without guiding the students too much, since it is difficult to manage so many students in the classroom (...) Sometimes I felt I wanted to persist in supporting one group but allowing them to think [by themselves] and to find the error (...) but sometimes it was impossible since I had many groups [in the class] which were dependent from my help.

Also, the planned moment for shared discussions did not always meet the teachers' expectations. Some teachers considered that leading whole class discussions was quite challenging for them due to factors such as the time restrictions and classroom physical conditions that limited the involvement of all students in a conversation. As stated by Andreia: "The biggest difficulty was managing the class dynamics according with the existing physical conditions, calling all students' attention so they could hear one another and they could concentrate on the requested work and on the discussions". Some teachers recognized that the success of whole class discussions also requires that they develop specific actions even before that moment. For instance, to guarantee that students are able to participate in a significant way, a small number of teachers referred to the importance of constantly reminding the students to write down their answers. As commented by Isabel, there is "the need for constant monitoring of students' written records so that, at the moment of collective discussion, students are able to present the justifications and to sustain the strategies they used".

In that moment it may not be easy to call on all the different representations made by the students in the computer which, for instance, Alice considers to be a consequence of her lack of knowledge about the software and "of specific knowledge to teach statistics with a focus on statistical reasoning". However the majority of the teachers involved in the DRP do not mention these aspects in their answers to the questionnaire (Fig. 14.4).

14.6 Conclusions

Our study focused on the teachers' perspectives regarding the opportunities to promote students' statistical reasoning using a specific technology environment as well as the challenges the teachers identified in that context. The main findings of the study are summarized and discussed below, taking into account the dimensions associated with three main elements of SRLE (Garfield and Ben-Zvi 2009) considered in the analysis: the tasks, the software and the teacher's role in promoting the classroom discourse.

First, regarding the teachers' perspectives about the potential of the learning environment to promote students' statistical reasoning, the teachers made a very positive assessment of what had been achieved within the DRP. They were aware of many important features of the tasks, the selected software and the way they were combined, with the goal of promoting statistical reasoning. In their teaching practice, the teachers were used to establish descriptive statistics as a main learning objective (Leavy 2010), which did not happen in these lessons. In fact, the teachers identified important statistical ideas and processes, namely the need for data, notions focused on distribution, such as variation and center, covariation, sampling and inference.

Despite the limited nature of the questionnaire (Fig. 14.4) that does not allow an in-depth reflection focused on all statistical ideas involved in the tasks they used, the teachers expressed their increasing awareness of specific aspects that characterize statistics, in particular statistical reasoning, thereby meeting what Groth (2008) argues as necessary to effectively implement the new curriculum guidelines. Nevertheless, the fact that they do not mention some important ideas of the SRLE model (Garfield and Ben-Zvi 2009), for example, the joint interpretation of a measure of central tendency with a dispersion measure, indicates the need to possibly make this aspect more visible when designing the instructional tasks and setting objectives with the teachers.

Regarding the integration of technology in these lessons, from the teachers' perspectives, students easily get familiar with TinkerPlots™, and its use helps them to focus on data analysis, allowing the creation of multiple graphical representations and the instant calculation of statistical measures. The teachers recognized that the visualization of different graphical representations, made easy by the software and encouraged by the tasks' questions, contributed to the development of central statistical ideas, such as distribution, variability and covariation. From their perspective, it also allowed the exploration of other specific aspects of the syllabus such as sampling, which is highly relevant in promoting the emergence of informal inferential practices. It was acknowledged that the use of real data associated with a meaningful context was facilitated by the software, which in turn enhanced the students' involvement in the work and led them to understand the need for data in decision-making. The teachers' perspective regarding the role of the software constitutes a relevant output of the DRP since they seem to fully understand its structural role in creating a learning environment that fosters statistical reasoning. This seems to be in opposition with some superficial uses of technology in the classroom just to comply with the syllabus (Groth 2008).

Additionally, all teachers mentioned their important role in promoting a classroom discourse that stands in sharp contrast with traditional teaching practices (Garfield and Ben-Zvi 2009). It is clear from their commentaries that they understand that these lessons should be organized in order to give students the opportunity to expand their autonomous work and discuss ideas with their peers. Therefore, teachers associate their non-directive role in these lessons with two main moments (students' work in the task and whole class discussion), considering that they need to be attentive to students' thinking, ask questions that promote their thinking and ask them to explain their ideas.

Second, an analysis of the teachers' perspectives about the challenges they face in creating the envisioned learning environment to promote students' statistical reasoning reveals that they recognize that the objectives established for these lessons are complex and the teaching practice is rather demanding. They are aware of students' difficulties, namely in making generalizations from samples, as they usually do not take data variability into account in the tendencies to be generalized, and they miss the use of probabilistic language when formulating those generalizations. These are issues that will need to be addressed in the future, when planning further interventions.

The use of technology in these lessons raised some challenges for teachers. Some of them believed they were not able to provide an adequate support to all groups due to the large number of students in the class and the diversity of representations they created with *TinkerPlots*TM. Given the reduced number of lessons and lack of previous experience, the teachers need to further develop their skills in using *TinkerPlots*TM in the classroom. However, the conditions required to use technology have also to be taken into consideration.

The challenges associated with promoting a distinctive classroom discourse that were mentioned by the teachers do not come as a surprise, since the majority of students in these classes were not acquainted with this kind of environment. This was a particularly sensitive aspect in this project due to the open nature of the tasks, the variety of representations that emerged from the students' activities and the required processes of argumentation, particularly in whole class discussions. The challenging aspects identified by the teachers stress the necessity of going deeper in anticipating the teacher's role for preparing students to fully participate in that part of the lesson (Menezes et al. 2015), which is a process that takes time, both for students and teachers. A detailed analysis of the teachers' practice in these lessons would give us important information regarding the specific difficulties they faced.

Due to the particular conditions created in the DRP, every teacher could count on another fellow teacher in each lesson to assist the students, which contributed to a sense of confidence on the part of the teachers to conduct these lessons. In fact, the benefits of having teachers participating in collaborative working environments with other teachers and researchers when innovative approaches are proposed have been documented in research (Goodchild 2014; Makar and Fielding-Wells 2011; Potari et al. 2010). Finally, assuming that sustained changes in teaching practices rely on individual and social processes of sense-making (März and Kelchtermans 2013) and considering the teachers' claim about the need to further their professional development for working under the SRLE perspective, it would be beneficial to continue supporting teachers with more initiatives and projects of a collaborative nature.

Acknowledgements This research was developed in the context of the *Project Developing Statistical Literacy: Student learning and teacher education* (PTDC/CPE-CED/117933/2010) supported by *Fundação para a Ciência e Tecnologia*, in Portugal. Part of the data have been used in Henriques and Oliveira (2014), in Portuguese. We thank Ana Isabel Mota for the work carried out in data collection.

References

- Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, 13, 5–26.
- Batanero, C., Burrill, G., & Reading, C. (2011). *Teaching statistics in school mathematics—Challenges for teaching and teacher education: A joint ICMI/IASE study*. Dordrecht: Springer.
- Ben-Zvi, D. (2006). Using Tinkerplots to scaffold students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education—Proceedings*

- of the Seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.
- Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM Mathematics Education*, *44*, 913–925.
- Ben-Zvi, D., & Garfield, J. (2004). *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Ben-Zvi, D., & Makar, K. (2016). *The teaching and learning of statistics: International perspectives*. Switzerland: Springer.
- Chick, H. L., & Pierce, R. U. (2011). Teaching statistics at the primary school level: Beliefs, affordances, and pedagogical content knowledge. 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. 151–162). Dordrecht, The Netherlands: Springer.
- Cobb, P., Zhao, Q., & Dean, C. (2009). Conducting design experiments to support teachers' learning: A reflection from the field. *The Journal of the Learning Sciences*, *18*, 165–199.
- Denzin, N., & Lincoln, Y. (2005). Introduction: The discipline and practice of qualitative research. In N. Denzin & Y. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 1–19). Thousand Oaks, CA: Sage.
- Estrada, A., Batanero, C., & Lancaster, S. (2011). Teachers' attitudes toward statistics. 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. 163–174). Dordrecht, The Netherlands: Springer.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, R., et al. (2007). *Guidelines for assessment and instruction in statistics education: A pre-K-12 curriculum framework*. Alexandria, VA: The American Statistical Association.
- Garfield, J. B., & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics*, *31*(3), 72–77.
- Goodchild, S. (2014). Mathematics teaching development: Learning from developmental research in Norway. *ZDM Mathematics Education*, *46*, 305–316.
- Groth, R. (2008). Assessing teachers' discourse about the pre-K–12 guidelines for assessment and instruction in statistics education (GAISE). *Statistics Educational Research Journal*, *7*(1), 16–39.
- Henriques, A., & Oliveira, H. (2014). Ambientes de aprendizagem para promover o raciocínio estatístico dos alunos: O contributo das tarefas e da tecnologia [Learning environments to promote students' statistical reasoning: The contribution of the tasks and technology]. *Revista Educação Matemática em Foco*, *3*(2), 11–38.
- Henriques, A., & Oliveira, H. (2016). Students' expressions of uncertainty in making informal inference when engaged in a statistical investigation using Tinkerplots. *Statistics Educational Research Journal*, *15*(2), 62–80.
- Henriques, A. C., & Oliveira, H. (2013). Prospective teacher's statistical knowledge for teaching when analyzing classroom episodes. In A. M. Lindmeier, & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 41–48). Kiel, Germany: PME.
- Konold, C., & Miller, C. D. (2005). *TinkerPots: Dynamic data exploration* (version 1.1). [Computer software]. Emeryville, CA: Key Curriculum Press.
- Leavy, A. M. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning. *Statistics Educational Research Journal*, *9*(1), 46–67.
- 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., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. 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. 347–358). New York, NY: Springer.
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105.

- Martins, J. A., Nascimento, M. M., & Estrada, A. (2012). Looking back over their shoulders: A qualitative analysis of Portuguese teachers' attitudes towards statistics. *Statistics Educational Research Journal*, 11(2), 26–44.
- März, V., & Kelchtermans, G. (2013). Sense-making and structure in teachers' reception of educational reform. A case study on statistics in the mathematics curriculum. *Teaching and Teacher Education*, 29, 13–24.
- ME (2007). *Programa de Matemática do Ensino Básico [Basic Education Mathematics Syllabus]*. Lisboa: DGIDC.
- Menezes, L., Oliveira, H., & Canavaro, A. P. (2015). Inquiry-based teaching: The case of Célia. In U. Gellert, J. Gimenez Rodrigues, C. Hahn, & S. Kafoussi (Eds.), *Educational paths to mathematics* (pp. 305–321). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-15410-7_20.
- Oliveira, H., & Henriques, A. C. (2015). Characterizing one teacher's participation in a developmental research project. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the 9th Congress of the European Society for Research in Mathematics Education* (pp. 2881–2887). Prague: Charles University in Prague, Faculty of Education and ERME.
- Paparistodemou, E., & Meletiou-Mavrotheris, M. (2008). Developing young children's informal inference skills in data analysis. *Statistics Education Research Journal*, 7(2), 83–106.
- Potari, D., Sakonidis, H., Chatzigoula, R., & Manaridis, A. (2010). Teachers' and researchers' collaboration in analysing mathematics teaching: A context for teacher reflection and development. *Journal of Mathematics Teacher Education*, 13, 473–485.
- Watson, J. (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. (2008). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal*, 7(2), 59–82.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–265.
- Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58.

Chapter 15

A Study of Indonesian Pre-service English as a Foreign Language Teachers Values on Learning Statistics



Khairiani Idris and Kai-Lin Yang

Abstract Pre-service English as a Foreign Language (EFL) teachers are service students of a statistics course who will apply statistics knowledge as a tool in their future profession. Their future learning of statistics might be related to the value they have for statistics at the end of the course. By using a phenomenographic approach, we investigated 38 Indonesian pre-service EFL teachers' values on learning statistics. Three components of values on learning statistics were identified, which can be related to the components from task-value theory: intrinsic, attainment, and utility. The participants could be categorized as having either positive or negative values for each component. In addition, some conflicting characteristics were noticed, which could reflect the characteristics of Indonesian pre-service EFL teachers. Implications for college statistics teaching and future research are discussed.

Keywords Indonesian pre-service EFL teachers · Introductory statistics
Values on learning statistics

15.1 Introduction

The goal of college statistics courses is to produce statistically educated students, meaning they should develop statistical literacy (Aliaga et al. 2005; Franklin and Garfield 2006), which is a fundamental skill required by citizens (Rumsey 2002; Utts 2003) and a way of thinking in society (Giesbrecht 1996). For most service students in introductory statistics courses, statistics is an essential tool for their future profession (Scheaffer and Stasny 2004) as well as for doing their undergraduate research, which is a part of study requirements in some countries, like Indonesia (Sailah 2014).

K. Idris (✉)

State Institute for Islamic Studies of Lhokseumawe, Aceh Province, Indonesia

e-mail: khairiani@iainlhokseumawe.ac.id

K.-L. Yang

National Taiwan Normal University, Taipei, Taiwan

e-mail: kailin@ntnu.edu.tw

© Springer Nature Switzerland AG 2019

G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_15

Pre-service English as a Foreign Language (EFL) teachers, in particular, have at least three professional roles in the future in which they might need to utilize statistics. First, as future teachers, they might need statistical investigation skills for doing educational research including analyzing their own teaching which necessitate learning investigation process (Heaton and Mickelson 2002). Learning the investigation process also entails a deep understanding of statistical knowledge, not only about how to carry out data analysis, but also about how to find problems and formulate investigation questions (Franklin and Garfield 2006; Wild and Pfannkuch 1999). Second, as English Language Learners, pre-service EFL teachers may also require statistics to improve their knowledge and expertise. For example, since most of the useful studies published in linguistics journals are quantitative in nature (Lazaraton 2000), they would need statistical skills to understand findings from these studies. Third, statistics might be used as one of the contents they would introduce in their reading classes, because the 'content' in teaching language may go beyond literature. Moreover, due to the emergence of English immersion programs in some countries to promote students' English language competencies while learning content subjects, including mathematics (e.g., Cheng et al. 2010; Padilla and Gonzalez 2001; Yushau 2009), English language teachers may play a crucial role in such program. Referring to the suggestion from Crandall (1987), EFL teachers may use the relevant mathematics texts that students will read in their mathematics classes, which is significant to enable students to acquire reading skills they can apply in learning mathematics. Thus, learning statistics, which is one of the contents included in school mathematics, is one way to promote EFL teachers' knowledge and expertise which has been identified as increasingly marginalized within content areas (Cross 2011). Due to the evolving nature of statistics as a discipline, the knowledge should be subject to lifelong learning for pre-service EFL teachers. Moreover, when reading is viewed as an approach to learning content, not only content area teachers' perceptions and values on the content area reading (Hall 2005) but also language teachers' values about the content are required for investigation in the relevant research.

Several studies have shown that students' further involvement in the activities related to the subject learned is associated with the way they value the learning of the subject (Eccles and Wigfield 2002; Liem et al. 2008; Yang 1999). Likewise, students' appreciation of the course was found to dominate their willingness to learn statistics and led to a mature approach in learning the course (Gordon 1995; Petocz and Reid 2005). This may imply the need for pre-service EFL teachers to hold a disposition towards the values on the learning of statistics in addition to the values on the knowledge obtained through investigation (Heaton and Mickelson 2002) because they might need to learn and utilize statistical investigation in the future.

Despite the significance of students' values on learning, studies on this area are sparse in statistics education literature. Therefore, this study aimed to explore qualitatively different ways in which pre-service EFL teachers value the learning of statistics. Accordingly, the research questions proposed in this study were: How do pre-service EFL teachers value the learning of statistics and what different categories can be assigned to describe their values on learning statistics?

Task value (Eccles et al. 1983; Eccles and Wigfield 2002) is defined as the reasons or incentives students believe they would receive from engaging in the activity. We considered the theory would be able to describe the ways in which students' valued learning of statistics in this study when the task was specifically related to learning statistics. More elaboration of task value theory is discussed in the subsequent section.

15.2 Theoretical Framework

15.2.1 *Components of Task Value*

The word *value* in the phrase *task value* was defined by Eccles and Wigfield (2002) as the incentives or reasons for doing an activity (or task). Theorists have offered a variety of definitions of task value (e.g., Atkinson 1964; Eccles et al. 1985). These definitions, however, have a common theme: value can be the incentives, rewards, and/or attainment that one expects to obtain by engaging in a task. When students value something, they are more likely to engage in that behavior (Barron and Hulleman 2015). Furthermore, Eccles et al. (1985) proposed four different components of task values as reasons why a task would hold value for an individual: intrinsic value, attainment value or importance, utility value or usefulness of the task, and cost. The answer to a question like "Why do you want to do the task?" can capture the essence of this value the student has for a task (Wigfield and Eccles 2000) as described in more detail below.

Intrinsic value, or interest value, is the inherent enjoyment of pleasure one gets from engaging in a task for its own sake (Eccles et al. 1983; Eccles and Wigfield 2002). This includes a statement like "In general, I find working on math assignments is very interesting".

Attainment value is the importance of doing well on a task in terms of one's self-schema and core personal values (Eccles et al. 1983; Eccles and Wigfield 2002). That is, it reflects that the task affirms a valued aspect of an individual's identity and meets a need that is important to an individual such as to fulfill achievement or social needs. A statement such as "I feel that getting a high score in mathematics is very important for me," can be categorized in this component.

Utility value is defined as how well a task relates to current and future goal (Eccles et al. 1983; Eccles and Wigfield 2002). That is, it reflects the usefulness of the task for achieving short-term or long-term goals. This value is the extrinsic value, which was also labeled to emphasize engaging in a task as a means for achieving another end (Barron and Hulleman 2015). For example, a statement like "What I learn from mathematics class is useless for my daily life outside school" can be designated into this component.

Finally, cost is conceptualized in terms of negative aspects of engaging in the task. This component includes the amount of effort, time lost, and suffered as a consequence of engaging in the task. For example, the statement "doing statistics

task takes up too much time” can be included as the effort cost. Thus, the perceived costs associated with performing a task can negatively impact the overall value of the task (Eccles et al. 1983; Eccles and Wigfield 2002; Flake et al. 2015).

The theory of task value along with its components formed the basis for research design of this study. We began by designing open-ended questions as the means to capture the essence of pre-service EFL teachers’ perceived values for learning statistics. Afterwards, the components of task value were employed to interpret their responses as well as in generating categories for describing the qualitatively different ways in which they valued the learning of statistics.

15.2.2 Factors Related to Values on Learning Statistics

A general motivation for individuals to perform their best is the need for achievement (Atkinson 1964). An expectancy-value theory of achievement motivation developed later (Eccles et al. 1985; Wigfield and Eccles 2000) suggested that task values directly influence performance, persistence, and choice. Accordingly, Bandura (1989) stated that individuals’ actions reflect their value preference. This may imply that the way students value their learning would influence their learning behavior. Some empirical studies have also suggested that task value was related to the utilization of learning strategies and thus influences achievement (e.g., Pintrich and Schrauben 1992).

From the perspective of approaches to learning (Biggs 1985; Marton and Säljö 1984), it has been acknowledged that students adopt the learning strategy most appropriate for their motives, including their perceptions of the relevance of the subject for their study field (Lucas 2001). Students who were driven by extrinsic motivation would apply a surface learning strategy by memorizing facts from the books or lectures. Those who learned due to their intrinsic interest would apply deep learning strategy by reading widely and interrelating with previous knowledge. Students who learned due to the intention to attain a high grade would apply a strategic strategy by learning through a systematic way and schedule time to compete with other learners. From the description of task value components in the previous sub-section, we may relate intrinsic value component to intrinsic motive, utility value component to extrinsic motive, and attainment value component to attainment motive. Hence, there might be different learning strategies used by students in the different value components.

To date, studies on students’ values on learning statistics are sparse in literature. Nonetheless, some other constructs, such as attitudes toward statistics (e.g., Dauphinee et al. 1997; Schau et al. 1995; Wise 1985), can be highly related to values on learning statistics. In fact, values and attitudes are the two constructs that have been acknowledged to have a close relationship (Bishop et al. 1999). Accordingly, several components of attitudes toward statistics, such as beliefs about competence and about the difficulty of statistics, are among factors assumed to influence values (Eccles and Wigfield 2002). Some items used to measure attitudes toward statistics in SATS (Gal et al. 1997), such as “I will have no application for statistics in my

profession” (beliefs about the usefulness and relevance of statistics) and “I like statistics” (interest in learning statistics), are related to values on learning statistics. The former is related to utility value while the latter is related to intrinsic value. However, using such Likert-type items may limit the interpretability of the values expressed by students. Hence, in this study we would like to deepen our understanding regarding the different reasons students value the learning of statistics by using open-ended questions that can provide sufficient opportunity for students to express their values on learning statistics. Moreover, by applying the qualitative method we could reveal not only how students could be classified into different value components but also how students’ values on learning statistics were contrasted across the components.

On the other hand, the ways students value the learning of statistics may also be related to their conceptions of statistics, i.e., the extent to which they understand statistics. For instance, Petocz and Reid (2005) found that students with the limited conceptions of statistics perceived that there would not be any role for statistics in their future profession, while those who had the broader conceptions could see statistics as an important skill for their future profession. The perceived usefulness of statistics in future profession could be accounted as the utility value of learning statistics (Eccles and Wigfield 2002).

15.3 Method

Because values on learning may be highly related to cultures, we applied a phenomenographic method in this study to reveal pre-service EFL teachers’ value on the learning of statistics. The method, which was initiated by Marton and his colleagues (1981) in Sweden, was designed to answer particular questions about people’s thinking and learning. More specifically, Marton (1994) defined phenomenography as

the empirical study of the limited number of qualitatively different ways in which various phenomena in, and aspects of, the world around us are experienced, conceptualized, understood, perceived and apprehended. (p. 4424)

This research method has been used widely in educational research to identify different ways of experiencing teaching and learning (e.g., Crawford et al. 1998; Tsai 2004) as well as academic disciplines included in the area of statistics education (e.g., Gordon 2004; Idris and Yang 2017; Petocz and Reid 2005; Reid and Petocz 2002; Yang 2014).

15.3.1 Participants

Thirty-eight pre-service EFL teachers (29 females and 9 males) who were taking an introductory statistics course completed an open-ended questionnaire. These partic-

ipants will be referred to as ‘students’ in this chapter. The students were from the English Education Department of an Islamic college in Aceh Province, in the western part of Indonesia. The introductory statistics course is one of the compulsory courses for students in the English Education Department and is offered during their second year at the college. The course, which involves topics such as data displays, numerical summaries, examining data distributions, correlation analysis, simple linear regression, t-test and chi-square test, is expected to provide basic statistical knowledge and skills for students, before they take the Educational Research Methodology course as a preparation to conduct their undergraduate research at the end of their college education. The course consists of one lecture per week of 100-minute duration. The class is usually taught using a traditional lecture style, and the assessment includes individual and group assignments and quizzes as well as a midterm and final examination. In several class meetings, students are asked to take their personal portable computer to the class for learning data analysis with Microsoft Excel and SPSS software. This research was conducted around mid-semester, by which time students have learned the topics of data displays, numerical summaries, and examining data distributions.

Before taking the introductory statistics course, the students had learned some basic statistics as part of high school mathematics. The educational backgrounds of students in this study, however, varied, particularly related to experience and expertise in statistics, because the students came from different types of senior high schools (see Table 15.1) with a different amount of statistics content offered in their schools. General and Islamic senior high schools have similar mathematics curriculums, which differ from those in vocational schools. The latter provide less statistics content than do the former. In the Indonesian senior high school curriculum developed by *Badan Standar Nasional Pendidikan* (Office of National Standard for Education), basic statistics content is included in grade 11 and is limited to data displays and descriptive statistics such as measures of center and variability. There are also two content areas in basic probability: sampling space and probability of an event (Badan Standar Nasional Pendidikan [BSNP] 2006). Vocational schools have similar statistics content but with more limited hours of teaching. Less teaching of statistics content is also found in several majors of vocational school such as art, tourism, and household technology, which do not include probability at all (BSNP 2006).

15.3.2 Data Collection

The data in this study included 38 written responses and 23 interview transcripts, which were obtained from a two-stage data collection procedure. In the first stage, 38 students gave their written responses to three open-ended questions designed to investigate their values of learning statistics.

Q1. Do you study hard in the statistics course?

Table 15.1 Demographics of 38 students in questionnaire survey

| Types of senior high school attended | Statistical topics learned | | | Female | Male | Total |
|---|----------------------------|------------------------|---------------------|--------|------|-------|
| | Data displays | Descriptive statistics | Probability | | | |
| General senior high school ^a | ✓ | ✓ | ✓ | 17 | 4 | 21 |
| Islamic senior high school ^a | ✓ | ✓ | ✓ | 8 | 5 | 13 |
| Vocational high school ^b | ✓ | ✓ | majors ^c | 4 | 0 | 4 |
| Total | | | | 29 | 9 | 38 |

^aScience majors learned statistics for more hours with more advanced content compared to social and language majors

^bStatistics content and teaching hours were more limited than in the other schools

^cMajors: students majoring in art, tourism, and household technology did not learn this topic

Q2. Please give your reason

Q3. What targets do you set to achieve when learning statistics?

We used Q1, which asked about the intensity of learning activity, to capture students' engagement in learning statistics by considering that students are more likely to engage in doing something when they value it (Barron and Hulleman 2015). Afterwards, as there are many reasons why the task would hold value for a student (Wigfield and Eccles 2000), Q2 and Q3 were designed to capture the reasons students have for engaging in the learning of statistics. During the data collection we ascertained that students were informed that their participation would not affect their course performance, and they were allowed to choose whether to participate or not in this research.

After the administration of the questionnaire, 23 (19 females and 4 males) of the 38 students whose types of high schools' background were varied (i.e., 2 vocational, 7 Islamic, and 14 general senior high schools), volunteered to take part in personal interviews with a researcher. These interviews represent the second stage of the data collection. Each interview session, lasting from 30 min to one hour (median = 42 min), was situated to allow students to express and share their views conveniently. The interviewer brought along students' written responses and began each session by asking students to read their statements to the open-ended survey questions and then clearly elucidated their responses. Students were asked to give examples and probed to explain their written responses. For instances, probing questions used to delve into students' reasons related to their interest in learning statistics were "what factors make you like learning statistics?" and "what factors make you dislike learning statistics?"

The interviews were audio-taped and transcribed for further analysis. Follow-up interviews through internet social media and phone were also undertaken whenever possible for some aspects of the transcripts that needed clarification. On occasion, the lecturer teaching the introductory statistics course was involved in discussing the interview responses and her teaching of the course. Some documents, such as teaching materials, tests, and results, were also recorded as supporting data. For example, teaching materials and test items could uncover the source of confusion in the thinking of those students who dislike numerical activities.

15.3.3 Data Analysis

Analysis of written responses and interview transcripts was conducted with the aim of describing the ways in which students valued the learning of statistics. We kept this goal in mind during the data analysis in order to ensure the research questions of this study could be answered properly. As the data were in Indonesian language, two Indonesian native-speaker researchers took part in the analysis to explore the variations that emerged from the data. At the end of analysis, one researcher translated the categories and quotes from Indonesian to English and discussed it with a non-Indonesian speaking researcher to ensure the proper terms had been used in the translation to reflect the precise meanings.

This analysis took place in three main stages. In the first stage, we tried to capture the general features of each of the students regarding their values associated with learning statistics. We did this by repeatedly reading each of the written responses and interview transcriptions. These general features were then summarized for each student. The second stage was for identifying keywords to represent each feature found from the previous stage. The keywords were then grouped into different categories based on their similar characteristics. We further interpreted each keyword as an intrinsic value, attainment value, utility value, or cost component and then separated the negative and positive responses within each component. The data analyses in these first two stages were carried out independently by the two researchers.

In the next stage, comparisons and discussions were used to find agreement with the initial draft of categories. In this process, re-examining the data was carried out together at some places to assure that all information was properly interpreted. The classification of students into each of the finalized categories was recorded upon agreement of both researchers. Several corresponding keywords for each value component were selected and are shown in the second columns of Table 15.2.

Table 15.2 Components of values on learning of statistics

| Components | | Keywords | No. of responses (%) | Sample quotes |
|-------------------|-------------|--|----------------------|--|
| Intrinsic values | (+) | Interested, challenged, curious, satisfied | 22 (57.9%) | <i>I feel satisfied every time I can get the correct solutions</i> |
| | Both | | 6 (15.8%) | |
| | (-) | Dislike, bored, confused, complicated | 10 (26.3%) | <i>Too many confusing formulas</i> |
| Attainment values | Beyond | | 2 (5.2%) | |
| | (+) | High grade, outperform, excel | 18 (47.4%) | <i>The important thing is I can get A+</i> |
| | (-) | Pass the course, pass the exam | 9 (23.7%) | <i>I study statistics to pass the exam</i> |
| | No evidence | | 9 (23.7%) | |
| Utility values | (+) | Thesis, career, daily life | 34 (89.5%) | <i>It'll be helpful in doing my thesis research</i> |
| | (-) | Useless, no benefit | 2 (5.3%) | <i>I haven't seen any benefit from what I've learned</i> |
| | No evidence | | 2 (5.3%) | |

15.4 Findings

15.4.1 Three Components of Values on Learning Statistics

We found three rather than four components of values on learning statistics in this study: intrinsic, attainment, and utility values. The cost component was not identified in our data, which might be due to the limitation of methodology used, particularly related to the proper questions proposed in interviews to probe student thinking of this component. Nonetheless, we may argue that empirically, cost was not the component that would be thought naturally by our students in expressing their value on learning statistics. Moreover, the quantitative empirical research reported in literature (e.g., Eccles and Wigfield 1995; Parsons et al. 1984) has also suggested that task value could be represented by the three components. Despite the different methodology used, our study could echo the findings of these studies.

Table 15.2 presents the keywords used in categorizing responses into the corresponding components, and the number of responses for each component, which is also shown as a percentage of 38 students, to indicate the proportion of respondents.

Brief sample quotes retrieved from interview transcripts or written responses are also provided as the illustrative features of the corresponding groups.

We initiated the analysis by attempting to classify each student into either positive or negative values for each component based on his/her responses for the open-ended questions and/or interview. Yet, we found that some written responses had insufficient information about the attainment and utility components. We classified those students into “no evidence” in the two components.

Intrinsic values, which correspond to the enjoyment students get from learning statistics, could be identified from all students’ responses. 22 students expressed positive intrinsic value, while 10 students expressed negative intrinsic value on learning statistics. On the other hand, we found six students expressing both positive and negative responses, which could not be interpreted precisely as positive or negative value. For example, S5 stated his deep satisfaction in learning statistics:

The feeling when getting correct solutions of difficult problems... it’s like the feeling when achieving a big success...

However, in another part of the transcript, he gave another statement about his unfavorable feeling about statistics:

...because I have to memorize those confusing formulas that mix-up of alphabetical letters and numbers.

Similar cases also emerged from five other students, who had common characteristics of having low interest in statistics but sometimes feeling challenged and satisfied. Hence, we categorized these students as having both positive and negative values of intrinsic component.

As for attainment values, we defined a positive value as one representing students who express an ambition and set a target for obtaining a high grade or outperforming other students in learning statistics. The negative value, in contrast, represented a group of students with avoidance components of need-achievement motivation (Atkinson 1964) who set the lowest target in learning such as to pass the exam or the course. There were nine insufficient responses from this component which we classified as no evidence. In addition, we found two students expressed distinctive response related to attainment value which could neither be classified as positive nor negative attainment value in learning statistics. These students conveyed their view that they put more concern on understanding the materials rather than scores or competitive ambition. We quote one of these statements below:

...scores will follow when we can understand (the materials), the important thing is to understand. After understanding, the chance of making mistakes will certainly be lessen... I do not target the scores.

This expression might exhibit the views on attainment to be beyond scores, to gain mastery. Hence, we assigned the two students’ responses into the beyond attainment values component.

Utility values, on the other hand, could be determined for 36 students, two of whom could be classified easily into negative values. These two students were unaware about

the usefulness of learning statistics and claimed that the course was meaningless for them and they could not see any benefit in learning it. We also observed that most of students' responses within the positive utility values mentioned that one of the usefulness of statistics was for doing their undergraduate research.

15.4.2 Relationships Among Components

There was another finding that emerged from the analysis of values on learning statistics—a student having a positive value in one component was not an indicator that he/she was also positive in other components. Tables 15.3, 15.4 and 15.5 show, respectively, the number of students for the relationships between intrinsic and attainment, intrinsic and utility, and attainment and utility components. As for students who had beyond positive and negative attainment value, we noticed that they expressed positive values in both intrinsic and utility components (see Tables 15.3 and 15.4), which means that they could feel enjoyment in learning statistics and believed in the usefulness of learning the course for their future. Table 15.3 shows that, approximately, 32% (12 out of 38) of the students had positive values in both intrinsic and attainment values on learning statistics. There were about 10% (4 out of 38) of the students who expressed positive intrinsic and negative attainment values on learning statistics. One of them mentioned her low confidence or negative ability beliefs (Eccles et al. 1983) in statistics that made her set a low target in learning statistics.

Table 15.3 Relationship between intrinsic and attainment components

| | | Intrinsic | | | Total |
|------------|-------------|-----------|------|-----|-------|
| | | (+) | Both | (-) | |
| Attainment | Beyond | 2 | – | – | 2 |
| | (+) | 12 | 3 | 3 | 18 |
| | (-) | 4 | 1 | 4 | 9 |
| | No evidence | 4 | 2 | 2 | 9 |
| | Total | 22 | 6 | 10 | 38 |

Table 15.4 Relationship between utility and attainment components

| | | Utility | | | Total |
|------------|-------------|---------|-----|-------------|-------|
| | | (+) | (-) | No evidence | |
| Attainment | Beyond | 2 | – | – | 2 |
| | (+) | 17 | – | 1 | 18 |
| | (-) | 8 | – | 1 | 9 |
| | No evidence | 7 | 2 | – | 9 |
| | Total | 34 | 2 | 2 | 38 |

Table 15.5 Relationship between intrinsic and utility components

| | | Intrinsic | | | Total |
|---------|-------------|-----------|------|-----|-------|
| | | (+) | Both | (-) | |
| Utility | (+) | 21 | 6 | 7 | 34 |
| | (-) | – | – | 2 | 2 |
| | No evidence | 1 | – | 1 | 2 |
| | Total | 22 | 6 | 10 | 38 |

“No, I don’t target high score because I know my own ability, it seems that I’m not competent enough to get high scores in statistics, even though I like this course...”

On the contrary, about 8% (3 out of 38) of students expressed negative intrinsic and positive attainment values. This may simply explain that the students’ motive in learning was an *achieving motive* (Biggs 1985), with which they have high ambition for outperforming others in class whether or not they like the course. For example, one of these students stated her dislike of the statistics course in her written response.

“I don’t really like it. The materials are confusing and there are too many formulas.”

Yet, later she stated that to get the high score is her target in learning statistics.

“My target in learning this course is to obtain high scores.”

From Table 15.4 we may see that almost half of students (17 out of 38) showed positive values in both attainment and utility components. There were 8 students who expressed positive utility together with negative attainment values, which means that those students were aware about the usefulness of learning statistics, yet they did not think that it would be important to perform well in learning statistics.

We also found that, from Table 15.5, out of the 34 students within the positive utility value, 7 of them were grouped into the negative intrinsic value. A sample quote from one of students within this group is as below.

“Statistics will be helpful for doing my thesis research, but I don’t really like it. It’s very difficult and confusing.”

15.5 Discussions and Suggestions

15.5.1 Characteristics of Indonesian Pre-service EFL Teachers’ Value on Learning Statistics

In this study, we explicitly uncovered the ways in which Indonesian pre-service EFL teachers’ value their learning of statistics through the three components of task value defined by Eccles et al. (1983): intrinsic components, attainment component, and

utility component. Based on the interview and open-ended question responses, the students can be classified into either positive or negative value in each component. Particularly for intrinsic and attainment components, we found two categories of values on learning statistics that did not seem to have been reported in the prior studies related to task value, i.e., both positive and negative intrinsic and beyond positive and negative attainment values.

Among these components, the positive values of the utility component accounted for the highest numbers of students, particularly those who recognized the use of statistics for their undergraduate research. One reason for this finding might be due to the statistics lecturer who often emphasized the connection of the materials to students' undergraduate research for the purpose of motivating their learning, as revealed from the interview with the lecturer. One of characteristics of students from South-East Asian countries (including Singapore, Philippines, and Indonesia) is that they would tend not to challenge or disagree with what their teacher says in the classroom (Liem et al. 2009), even though they did not fully grasp the ideas. This phenomenon was echoed in our finding, in which we found that there were some students who believed that they need statistics since their lecturer had conveyed from the beginning of the course that it would be useful for their undergraduate research.

The expression of both positive and negative values on the intrinsic component, on the other hand, was evident in six students, who expressed their enjoyment in learning statistics as something challenging and giving personal satisfactions. At the same time, they also indicated their unfavorable feelings toward the course due to some specific parts such as confusing and dense formulas they need to use. These students can be categorized as holding conflicting conceptions or beliefs (Marton and Pong 2005). We suggest that such characteristics of students need to be taken into consideration when statistics lecturers design teaching activities or materials for students learning statistics. Emphasizing conceptual understanding rather than relying heavily on formulas is considered more effective for student learning in introductory statistics class (Aliaga et al. 2005; Franklin and Garfield 2006).

This conflicting beliefs characteristic is difficult to identify through closed questionnaire surveys since students would tend to respond based on what they believe is socially desirable (Marton and Pong 2005). Therefore, this might be one possible reason for the high attitude in learning mathematics of Indonesian students reported in international surveys. For instance, it would be more unlikely for these students to choose disagree in their response to a questionnaire statement like "I enjoy learning many interesting things in mathematics," because they did enjoy some parts in learning mathematics. Therefore, this finding could also shed light on how the items for value on learning statistics questionnaire should be developed. Since a student may value the procedural part of statistics differently from its conceptual part, we suggest that it would be more meaningful to combine the how and what to value in an item. For example, the items assessing intrinsic value in a questionnaire of task value such as "In general, I find working on math assignment very boring" (Eccles and Wigfield 1995) can be modified for statistics learning as: "I find working on the procedural part in statistics assignment very boring".

As for the value of beyond positive and negative attainment, the two students who expressed this value were found to hold positive value in the other two components. Since one of these students had an average score in statistics, we cannot assume that such value had resulted from their high performance. Those students can be characterized as mastery-oriented students (Ames 1992), because they will choose challenging tasks and were more concerned with their own progress than with outperforming others. Such kind of learning characteristic can be expected as the key to lifelong learning. Hence, we should put more concern on developing the value of beyond positive and negative attainment in learning statistics, as the initial step for preparing students to be lifelong learners. Moreover, referring to Biggs and Tang (2011), students will change from a surface to a deep approach to learning statistics only by knowing why and how statistics should be learned. Thus, one possible way for preparing students for their lifelong learning in statistics class is by directing them not only on why to learn but also on how to learn statistics.

15.5.2 Conflicting Values on Learning Statistics

The findings of this study also showed that some students hold conflicting values among the three components. That is, they expressed a negative value in one component while holding a positive value in another component. Since task value is shaped over time by individual and contextual factors which include past experiences and cultures (Eccles et al. 1983), the conflicting values found among Indonesian pre-service EFL teachers might be explained by these factors. For example, the discrepancy between intrinsic and attainment values might be the result of the culture of Indonesia as a collectivist country (Hofstede 2001). Students from collectivist societies are less competitive than those in individualist societies and prefer to work together as a group, and they would not mind being modest to suit and conform to the majority of students (Triandis 1995). Such characteristics might lead students to have less ambition to outperform others or to be the best in class.

On the other hand, students who expressed negative attainment together with a positive intrinsic value may indicate a low self-ability belief (Eccles and Wigfield 1995). These students tend to set the target as low as they can to just pass the course examination although they have positive intrinsic and utility values in learning statistics. This phenomenon could be investigated in future studies, as the interplay between ability beliefs has been shown to be positively related to effort and persistence (Wigfield and Eccles 2000) as well as performance (OECD 2014). Furthermore, in order to encourage students' belief that ability can be improved, statistics lecturers may provide less challenging problems in the initial stage—such as those with familiar contexts that can be partially solved based on students' prior knowledge. Additionally, collaborative learning environments, such as team discussions and game activities related to statistical principles (Davis and Blanchard 2004), are considered suitable for students from collectivist cultures, but competitive learning between groups can be used when students review learned materials

(Kolawole 2008). Setting up such learning environments are among the many ways to encourage students' self-ability belief, which can lead them towards the positive attainment values on learning statistics.

Furthermore, while almost 90% of the students expressed positive utility value, only 60% of them had a positive intrinsic value, which indicated the need for more concern about developing more positive intrinsic values towards learning statistics. This phenomenon may further reflect the characteristics of students in East-Asian countries, as discussed previously, i.e., they would not disagree with what their teachers said in the classroom even though they did not fully grasp the ideas. Thus, making students really view statistics as meaningful and useful knowledge that promotes their development may increase their positive intrinsic value. Various teaching and learning approaches have been suggested in literature to make students see the relevance of statistics for their life and future professions, such as doing the real statistics investigations (Smith 1998) and implementing farmer market projects for business students (Hiedemann and Jones 2010). Particularly for pre-service EFL teachers, we suggest that embedding the statistics contexts in English reading tests or using quantitative research articles from linguistics journals in teaching materials as one way to improve their interest in learning the course. Because valuing results from internalization and integration, which require that students are able to feel competent, related, and autonomous while doing the activities (Deci et al. 1991), the design of teaching may play an important role in developing values about learning statistics.

15.6 Conclusions

In summary, we have explored the different ways in which students value learning statistics on the basis of the motivation theory of task value (Eccles et al. 1983; Eccles and Wigfield 2002). This study sheds light on the elaboration of this theory, more particularly for statistics learning. Moreover, some critical value components were identified which revealed the variation of Indonesian pre-service EFL teachers' values about learning statistics. This variation, to some extent, might be explained by characteristics specific to Indonesian pre-service EFL teachers, since individuals' values on learning are shaped by their past experiences and social stereotypes (Eccles et al. 1983). The specific characteristics of Indonesian pre-service EFL teachers may include previous experiences in learning statistics and the role of statistical skills in their study field and future profession. Additionally, the conflicting values held by some students could be related to the aspect of Indonesian cultures. Thus, the findings of this study might be generalized to other service students in college statistics course who have some common characteristics with our students, such as having diverse experiences in learning statistics as part of school mathematics and the possibility to utilize the knowledge in their future professions. The conflicting beliefs found in intrinsic value components suggested considering the conditions of learning including what and how to learn as an additional information in the intrinsic component. Besides, the value of beyond attainment can be added in addition to

the items indicated positive or negative attainment value in the questionnaire of task value. For example, a statement like “In studying the course, I focus more on understanding the materials than just gaining high scores” may indicate beyond positive and negative attainment values.

Acknowledgements The development of this paper was supported by a grant from the Ministry of Science and Technology (MOST 104-2511-S-003-006-MY3). We are grateful for the ICME-13 TSG-15 team and participants for their valuable comments and suggestions toward improving the paper.

References

- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Utts, J., & Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education: College report* (Vol. 30). Alexandria, VA: American Statistical Association. Retrieved from <http://www.amstat.org/education/gaise/>.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261–271.
- Atkinson, J. W. (1964). *An introduction to motivation*. Oxford, England: Van Nostrand.
- Badan Standar Nasional Pendidikan. (2006). *Standar isi untuk satuan pendidikan dasar dan menengah*. [Content standard for primary and secondary education]. Jakarta: Badan Standar Nasional Pendidikan.
- Bandura, A. (1989). Social cognitive theory. In R. Vasta (Ed.), *Annals of child development: Vol. 6. Six theories of child development: Revised formulation and current issues* (pp. 1–60). Greenwich: JAI Press.
- Barron, K. E., & Hulleman, C. S. (2015). Expectancy-value-cost model of motivation. In J. D. Wright (Ed.), *International Encyclopedia of the social & behavioral science* (pp. 261–271). Oxford: Elsevier Ltd.
- Biggs, J. B. (1985). The role of metalearning in study processes. *British Journal of Educational Psychology*, 55(3), 185–212. <https://doi.org/10.1111/j.2044-8279.1985.tb02625.x>.
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does*. UK: McGraw-Hill Education.
- Bishop, A., FitzSimons, G., Seah, W. T., & Clarkson, P. (1999). *Values in mathematics education: Making values teaching explicit in the mathematics classroom*. Paper presented at 1999 Australian for Research in Education Annual Conference. Available online at <http://files.eric.ed.gov/fulltext/ED453075.pdf>.
- Cheng, L., Li, M., Kirby, J. R., Qiang, H., & Wade-Woolley, L. (2010). English language immersion and students’ academic achievement in English, Chinese and mathematics. *Evaluation & Research in Education*, 23(3), 151–169.
- Crandall, J. (1987). *ESL through content-area instruction: Mathematics, science, social studies. Language in Education: Theory and Practice*. NY: Prentice-Hall, Inc.
- Crawford, K., Gordon, S., Nicholas, J., & Prosser, M. (1998). University mathematics students’ conceptions of mathematics. *Studies in Higher Education*, 23(1), 87–94.
- Cross, R. (2011). Troubling literacy: Monolingual assumptions, multilingual contexts, and language teacher expertise. *Teachers and Teaching*, 17(4), 467–478. <https://doi.org/10.1080/13540602.2011.580522>.
- Dauphinee, T. L., Schau, C., & Stevens, J. J. (1997). Survey of attitudes toward statistics: Factor structure and factorial invariance for women and men. *Structural Equation Modeling: A Multidisciplinary Journal*, 4(2), 129–141.

- Davis, N. T., & Blanchard, M. R. (2004). Collaborative teams in a university statistics course: A case study of how differing value structures inhibit change. *School Science and Mathematics, 104*(6), 279–287.
- Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist, 26*(3–4), 325–346.
- Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., et al. (1985). Self-perceptions, task perceptions, socializing influences, and the decision to enroll in mathematics. In S. F. Chipman, L. R. Brush, & D. M. Wilson (Eds.), *Women and mathematics: Balancing the equation* (pp. 95–121). NY: Psychology Press.
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., & Meece, J. L. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: W.H. Freeman.
- Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*(3), 215–225. <https://doi.org/10.1177/0146167295213003>.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>.
- Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology, 41*, 232–244. <https://doi.org/10.1016/j.cedpsych.2015.03.002>.
- Franklin, C. A., & Garfield, J. B. (2006). The GAISE project: Developing statistics education guidelines for grades Pre-K-12 and college courses. In G. F. Burril & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance: 2006 NCTM yearbook* (pp. 345–376). Reston, VA: NCTM.
- Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37–51). Amsterdam, Netherlands: International Statistical Institute/IOS Press.
- Giesbrecht, N. (1996). Strategies for developing and delivering effective introductory-level statistics and methodology courses. *ERIC Document Reproduction Service*, No. 393–668, Alberta, BC. Retrieved from <http://eric.ed.gov/?id=ED393668>.
- Gordon, S. (1995). What counts for students studying statistics? *Higher Education Research and Development, 14*(2), 167–184. <https://doi.org/10.1080/07294369501402023>.
- Gordon, S. (2004). Understanding students' experiences of statistics in a service course. *Statistics Education Research Journal, 3*(1), 40–59.
- Hall, L. A. (2005). Teachers and content area reading: Attitudes, beliefs and change. *Teaching and Teacher Education, 21*(4), 403–414.
- 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.
- Hiedemann, B., & Jones, S. M. (2010). Learning statistics at the farmers market? A comparison of academic service learning and case studies in an introductory statistics course. *Journal of Statistics Education, 18*(3). Available online at www.amstat.org/publications/jse/v18n3/hiedemann.pdf.
- Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks, CA: Sage Publications Ltd.
- Idris, K., & Yang, K. L. (2017). Development and validation of an instrument to measure Indonesian pre-service teachers' conceptions of statistics. *The Asia-Pacific Education Researcher, 26*(5), 281–290. <https://doi.org/10.1007/s40299-017-0348-z>
- Kolawole, E. B. (2008). Effects of competitive and cooperative learning strategies on academic performance of Nigerian students in mathematics. *Educational Research and Reviews, 3*(1), 33–37.
- Lazaraton, A. (2000). Current trends in research methodology and statistics in applied linguistics. *TESOL Quarterly, 34*(1), 175–181.

- Liem, A. D., Lau, S., & Nie, Y. (2008). The role of self-efficacy, task value, and achievement goals in predicting learning strategies, task disengagement, peer relationship, and achievement outcome. *Contemporary Educational Psychology, 33*(4), 486–512.
- Liem, G. A. D., Martin, A. J., Nair, E., Bernardo, A. B. I., & Prasetya, P. H. (2009). Cultural factors relevant to secondary school students in Australia, Singapore, the Philippines and Indonesia: Relative differences and congruencies. *Australian Journal of Guidance and Counselling, 19*(2), 161–178. <https://doi.org/10.1375/ajgc.19.2.161>.
- Lucas, U. (2001). Deep and surface approaches to learning within introductory accounting: A phenomenographic study. *Accounting Education, 10*(2), 161–184.
- Marton, F. (1981). Phenomenography—describing conceptions of the world around us. *Instructional Science, 10*(2), 177–200.
- Marton, F. (1994). Phenomenography. In T. Husén & N. Postlethwaite (Eds.), *International Encyclopedia of education*. Oxford, England: Pergamon.
- Marton, F., & Pong, W. Y. (2005). On the unit of description in phenomenography. *Higher Education Research & Development, 24*(4), 335–348.
- Marton, F., & Säljö, R. (1984). Approaches to learning. In F. Marton, D. J. Hounsell, & N. J. Entwistle (Eds.), *The experience of learning* (pp. 36–55). Edinburgh: Scottish Academic Press.
- OECD. (2014). *PISA 2012 results in focus: What 15-year-olds know what they can do with what they know*. OECD Publishing.
- Padilla, A. M., & Gonzalez, R. (2001). Academic performance of immigrant and US-born Mexican heritage students: Effects of schooling in Mexico and bilingual/English language instruction. *American Educational Research Journal, 38*(3), 727–742.
- Parsons, J. E., Adler, T., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology, 46*(1), 26–43.
- Petocz, P., & Reid, A. (2005). Something strange and useless: Service students' conceptions of statistics, learning statistics and using statistics in their future profession. *International Journal of Mathematical Education in Science and Technology, 36*(7), 789–800. <https://doi.org/10.1080/00207390500271503>.
- Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. H. Schunk & J. L. Meece (Eds.), *Student Perceptions in the Classroom* (pp. 149–183). NY: Lawrence Erlbaum Associates Inc.
- Reid, A., & Petocz, P. (2002). Students' conceptions of statistics: A phenomenographic study. *Journal of Statistics Education, 10*(2), 1–12.
- Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education, 10*(3), 6–13.
- Sailah, I. (2014). *Buku Panduan Kurikulum Pendidikan Tinggi [Curriculum guide book for higher education]*. Jakarta: Direktorat Jenderal Pendidikan Tinggi.
- Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. D. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement, 55*(5), 868–875.
- Scheaffer, R. L., & Stasny, E. A. (2004). The state of undergraduate education in statistics: A report from the CBMS 2000. *The American Statistician, 58*(4), 265–271.
- Smith, G. (1998). Learning statistics by doing statistics. *Journal of Statistics Education, 6*(3), 1–10.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.
- Tsai, C. (2004). Conceptions of learning science among high school students in Taiwan: A phenomenographic analysis. *International Journal of Science Education, 26*(14), 1733–1750. <https://doi.org/10.1080/0950069042000230776>.
- Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician, 57*(2), 74–79.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. <https://doi.org/10.1006/ceps.1999.1015>.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–248.

- Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement*, 45(2), 401–405.
- 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. <https://doi.org/10.1007/s10763-014-9519-z>.
- Yang, N.-D. (1999). The relationship between EFL learners' beliefs and learning strategy use. *System*, 27(4), 515–535.
- Yushau, B. (2009). Mathematics and language: Issues among bilingual Arabs in English medium universities. *International Journal of Mathematical Education in Science and Technology*, 40(7), 915–926.

Part V
Statistics Curriculum

Chapter 16

A MOOC for Adult Learners of Mathematics and Statistics: Tensions and Compromises in Design



Dave Pratt, Graham Griffiths, David Jennings and Seb Schmoller

Abstract There are many adults with low mathematical/statistical knowledge who would like to enhance that understanding. There are insufficient teachers to respond to the level of need and so innovative solutions must be found. In the UK, the Ufi Charitable Trust has funded a project to develop a free open online course to offer motivated adults access to powerful ideas. We reflect on the tensions and compromises that emerged during its design. More specifically, referring to data collected from users, we consider the challenge of developing resources that will support heterogeneous students from unknown backgrounds, who may have already been failed by the conventional educational system and who will have no interactive tutor support within this course.

Keywords Citizen Maths · Familiar situations · Learning · Powerful ideas Purpose · Utility

16.1 Adult Learners of Mathematics and Statistics

Open online courses (MOOCs) are sometimes posed as offering an educationally and cost-effective way of enabling adults new opportunities to improve their grasp of a particular subject, without needing to enroll on a face-to-face course, and at a much lower cost per learning outcome than for an equivalent taught course. MOOCs,

D. Pratt (✉) · G. Griffiths
UCL Institute of Education, London, UK
e-mail: david.pratt@ucl.ac.uk

G. Griffiths
e-mail: g.griffiths@ucl.ac.uk

D. Jennings · S. Schmoller
Independent Consultant and UCL Academic Visitor, London, UK
e-mail: david@alchemi.co.uk

S. Schmoller
e-mail: seb@schmoller.net

© Springer Nature Switzerland AG 2019
G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_16

for example those provided under the banners of FutureLearn, Coursera, EdX or Udacity, are typically at or approaching undergraduate degree level. But within the general adult population there are major gaps in knowledge at much lower levels. Mathematical and statistical knowledge is a case in point. Khan Academy (www.khanacademy.org) provides open online resources for learning mathematics, though these are closely tied to the traditional ‘schools’ curriculum.

A substantial proportion of the adult population (reported as 29% in BIS 2013, p. 38) have basic numeracy, but are not confident users of key mathematical ideas in life and work, at, in UK terms, Level 2.¹ Statistically literate behaviour is, according to Gal (2002), predicated on the activation of five knowledge bases, including mathematics, and a cluster of supporting dispositions. Statistical literacy, defined as “the ability to interpret, critically evaluate, and communicate about statistical information and messages” (Gal 2002, p. 1), is one critical but neglected skill area that needs to be addressed if adults are to become more informed citizens and recognise the value of statistics in making decisions (Wallman 1993).

In 2013, a consortium comprising Calderdale College, UCL’s Institute of Education, and the awarding organization OCR secured funding from the Ufi Charitable Trust to develop a free open online course for self-motivated adults who want to improve their appreciation of mathematics and statistics at Level 2. The course, being free to the user, needed to be sustainable at very low cost per learner. The course “Citizen Maths: Powerful Ideas in Action” (www.citizenmaths.com) has been developed over the last 30 months, through four main iterations, during which period several thousand people have signed up for it. Citizen Maths covers five powerful ideas in mathematics: proportion, measurement, pattern, representation, and uncertainty. The focus of this paper is on the tensions and compromises in the design of the sections of the course on representation (powerful because it underpins data, graphs, distributions, sampling and bias), and, to a lesser extent, on uncertainty (powerful because it underpins probability, risk, odds, large and small scale effects). We illustrate the design of the course through example activities drawn from representation and uncertainty. In fact, these activities speak to several of the five suggestions to work with learners in ways that are different from the typical instructional methods as proposed by Gal (2002). We explain those connections when reporting the example activities below.

As the course has developed, we have collected information to evaluate and inform each iteration of the resource. To review the challenges, the following data has been collected and will be used:

- Information about participants at sign up including their purposes in taking part
- Data about participant completion of each element of Citizen Maths

¹In the UK system, Level 2 is the standard expected of school students by age 16. BIS (2013 p. xxviii) notes that individuals “with skills below Level 2 may not be able to compare products and services for the best buy, or work out a household budget.” This describes a floor for level 2 but at the higher end, it is expected that individuals will be able to solve multi-stage problems. For data handling, this would mean dealing with descriptive rather than inferential statistics and calculating probabilities for combined events but not formalising algebraically as distributions.

- Participant views of the various activities via rating scales and additional qualitative comments.

In this paper, we discuss four design challenges often arising when the constraints of the course conditions competed with our pedagogic ideals. The challenges addressed are: how to offer resources to heterogeneous students; how to engage students without tutor support; how to promote meaningful mathematics; and how to assess. For each of the four challenges, we discuss the issues involved and evaluate our response in light of the data collected.

16.2 Challenge 1: How to Offer Resources to Heterogeneous Students

16.2.1 Discussion of Challenge 1

In a conventional teaching situation, the classroom teacher or tutor will often know the students in a quite personal way but, even if this is not the case, there will be a number of known characteristics amongst the students. In the case of our MOOC, little is known about the students. We anticipated that our students, as adults, will have been through elementary and high school mathematics, taught with a focus on the skills, methods and concepts set out in typical school curricula. Although research gives some general pointers to what typical learners know and understand at Level 2, further complexity was added to the design challenge because we could expect our learners to be heterogeneous in what, as experienced adults, they might bring to the course.

We therefore took an early view that there was little point in offering a course that focused on those same methods, skills and concepts in the system that had already failed our students. Instead of focusing hard on techniques when we knew little about what techniques the students might already have grasped, we aimed to focus on a few powerful ideas that might give our students some insight into how the discipline can ‘get stuff done’ for them (Pratt 2012). In this respect, we were influenced by Papert’s ideas on how students should learn to mathematize before following a formal course in mathematics (Papert 1972).

Each powerful idea in Citizen Maths is structured according to different ways in which it might be experienced, called ‘Powerful-Ideas-In-Action’ (PIAs). Each PIA consists of three or four activities, in which the student learns how the mathematical idea might be useful for them in their personal, social, occupational or scientific lives—the same contexts used by PISA (www.oecd.org/pisa/). Here are two examples of activities from each of the PIAs in ‘Representation’, which broadly speaking focuses on statistical literacy. Students meet the PIA, ‘interpreting data’, for example, in the contexts of: (i) how voting translates into seats in the Houses of Parliament and (ii) using statistics on violence and alcohol; both contexts can be considered societal. Students meet ‘interpreting charts’, for example, in the context of: (i) trends

in media communications and (ii) how your household income compares to the rest of the country; both contexts could both be regarded as personal. Students use statistics when ‘comparing groups’ in the context, for example, of: (i) how many people live in the same house in different regions of the country (societal) and (ii) of battery lifetimes for different usage of mobile phones (personal). Similarly, students work on the Uncertainty powerful idea through the PIAs, ‘making decisions’, ‘playing games’ and ‘creating or using simulations’ in personal, societal and occupational contexts.

It is worth noting that various powerful ideas present different challenges when attempting to create opportunities for the idea to be seen as useful in the students’ everyday or working life. How one interprets statistical data is highly dependent on the context whereas other areas of mathematics are often presented formally as if they were context-free. Cobb and Moore (1997) argued that, although mathematicians draw on context for motivation, their ultimate concern is with abstract patterns. In contrast, patterns in data analysis only have value according to how those patterns interweave with a contextual story line. Indeed, Wild and Pfannkuch (1999) have noted the fundamental importance of context in statistical thinking, when they depicted such thinking as emanating from the raw materials of statistical knowledge, information in data and context knowledge.

Our approach in Citizen Maths has been somewhat radical insofar as we have attempted to design activities that present proportion, measurement and pattern as contextually meaningful as those in representation and uncertainty; in so doing, we recognize that this aspect of the project has been especially challenging.

16.2.2 Evaluating Our Response to Challenge 1

Prior to registering to use Citizen Maths, participants are encouraged to undertake a pre-course self-assessment which is intended to make sure that the individuals who do take part have some understanding of the level of mathematics involved and the commitment needed. By the end of February 2017, there were just under 19,000 individuals who had completed a pre-course self-assessment of whom just under 10,000 had gone on to register for the course.

What can we say about these individuals?

The age profile of participants (see Fig. 16.1) shows a good spread across age groups with modal group being in their 30s. There was a smaller percentage of the participants under 20 although it should be noted that the sign-up process required individuals to be older than 16 and thus the available cohort is smaller than other age groups.

The gender split is weighted towards women (see Fig. 16.2). This is consistent with broader educational participation, for example Bosworth and Kersley (2015) noted that there were 53.9% females to 40% males involved in apprenticeship programmes in the academic year 2012/13.

Fig. 16.1 Age profile of participants in Citizen Maths

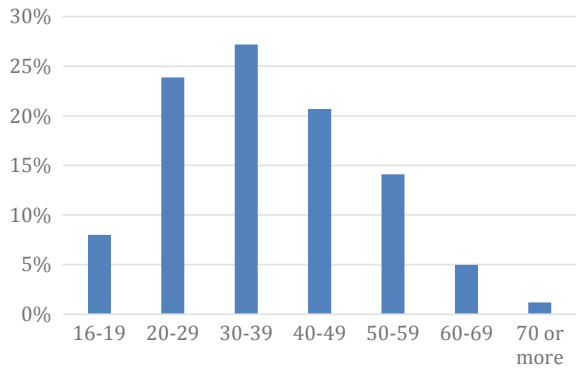
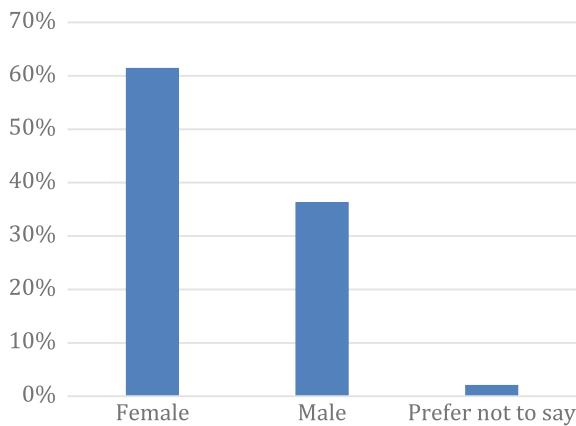


Fig. 16.2 Gender profile of participants in Citizen Maths



We asked the participants to identify the level of their highest qualification in any subject and for mathematics. The largest group did identify qualifications that were at NQF level 2 and below with significant numbers at Advanced level (level 3) and Higher education levels (level 4 and above). This includes those who have successfully progressed in non-mathematical subjects but have yet to achieve in mathematics.

We asked individuals to identify their reasons for taking part in Citizen Maths (see Fig. 16.3). It is worth noting that a good proportion of those signing up to Citizen Maths could be described as ‘interested professionals’ who have already succeeded at mathematics and have taken part in order to see how useful the resource is for others. Nevertheless, these are not a majority and as time moves on, and participant numbers increase, they are becoming an ever-smaller proportion of the total.

Overall, Citizen Maths has attracted males and females across ages who feel that they need to improve their mathematics and these individuals do have a range of other qualifications (although most have not achieved NVQ level 3). Whilst we do

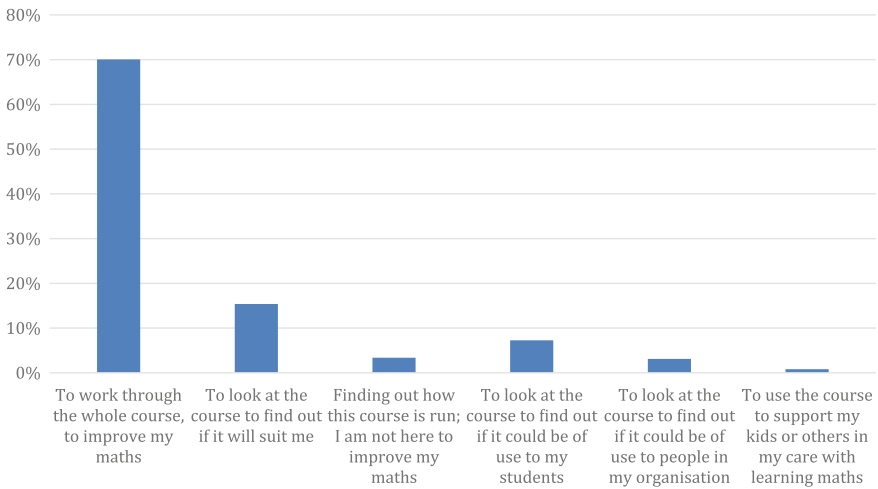


Fig. 16.3 The goals of the starters of Citizen Maths as selected when registering

have data about some other characteristics, it might be also be useful to know how participants are spread in relation to income or by profession. This might help us to understand for whom the course is most and least effective and to make changes. The problem is that the more questions that are asked, the larger is the barrier to participation.

We have noted some of the characteristics of those who have signed up to Citizen Maths and noted the mix. What is missing here is the extent to which the resource may or may not meet the needs of this group. In the evaluation of our response to Challenge 2 we summarise the feedback about the effectiveness of the course from users.

16.3 Challenge 2: How to Engage Students Without Tutor Support

16.3.1 Discussion of Challenge 2

Teachers in conventional classrooms are able to offer a personal level of interest and empathy with their students. In Citizen Maths there is no tutor present in real time. Moreover, the course aims to be sustainable in the future without the presence of a tutor. We decided to make the course feel personal by adopting techniques used by Peter Norvig and Sebastian Thrun in their very successful 2011 open online course “An introduction to artificial intelligence”. We used two ‘to-camera’ tutors, each of whom would introduce and develop particular PIAs. We make regular use of short

videos, sometimes involving the talking head of one of the tutors but often showing their hands or computer screen as they develop the mathematical idea in real time. Although no real-time interaction with students was possible, we hoped this would help to develop some intimacy, almost as if the tutor were talking directly with the student in their own home (Guo et al. 2014).

A limitation of this approach is that it is not possible to produce video that is equally engaging and correctly paced for a diverse set of students. The ‘feature’, however, that people can, if they wish, skip through videos or indeed repeat them may be an important benefit but with the associated risk that a student who skips through videos might miss a key learning point.

We recognized that these students already would have some initial motivation for joining the course and it was important to maintain engagement or eagerness that might be available at the outset. Face-to-face teachers might use their own personality to push through times when their class is less engaged. The best we could offer was to design purposeful tasks and this became very important in our approach in the light of other limitations.

Designing tasks that are seen as purposeful by the learners such that the learner comes to appreciate the power of the mathematics in that context is far from trivial, even in conventional classrooms (Ainley et al. 2006). Noss and Hoyles (1996) discuss what they call the ‘play paradox’; when a designer builds an environment that supports playful activity, the designer loses some control over what the learner might in fact learn. The teacher in a conventional classroom is unlikely to escape this same tension. With a clear obligation to a curriculum, teachers have to manage a corresponding ‘planning paradox’ (Ainley et al. 2006) as they attempt to inspire engagement without losing focus on that curriculum. This tension was alleviated to some extent in designing Citizen Maths by avoiding a highly prescriptive curriculum. We aligned ourselves to a philosophy in which the aim was to introduce students to the power of mathematics and statistics within their personal and social contexts. This felt less constraining than a commitment to a curriculum, especially one that might emphasize skills and techniques.

In discussing how to enhance statistical literacy, Gal (2002) proposed that a novel way to work with learners might be to focus on understanding results from polls, samples, and experiments as reported in the media. In fact, in one Citizen Maths activity, the student is required to predict the number of seats that a political party will gain, given the results of a prior opinion poll. The activity is first introduced by Noel-Ann, one of the two to-camera tutors, who describes the results of an opinion poll conducted on the day of the general election, and Noel-Ann then introduces an app, specially designed for the course. The app allocates seats randomly to the main political parties, according to probabilities set by the opinion poll results (see Fig. 16.4). The app introduces an element of playfulness as the student can run the app several times and perhaps note that the number of seats allocated varies though there is a limit to the extent of the variation. Utts (2003) has emphasised the importance of helping students to understand that variability is natural.

Of course, there is a danger that the student may miss this key learning point. In a conventional classroom, the teacher is able to assess the extent to which the

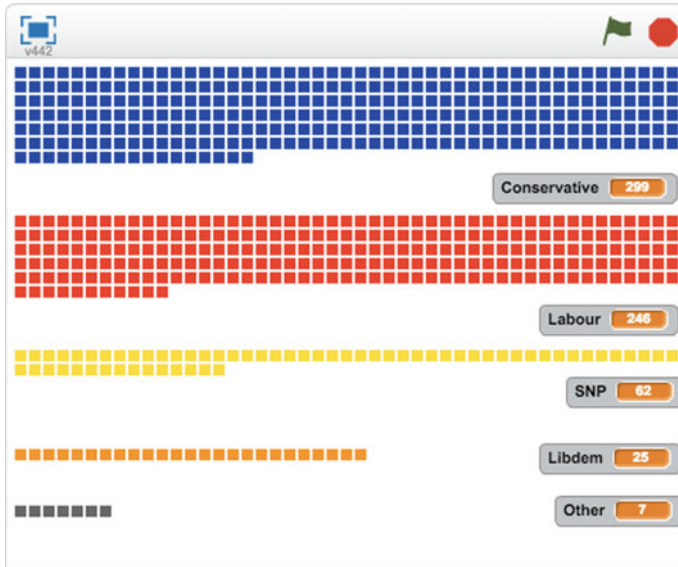


Fig. 16.4 The app allocates randomly seats to each of the main political parties

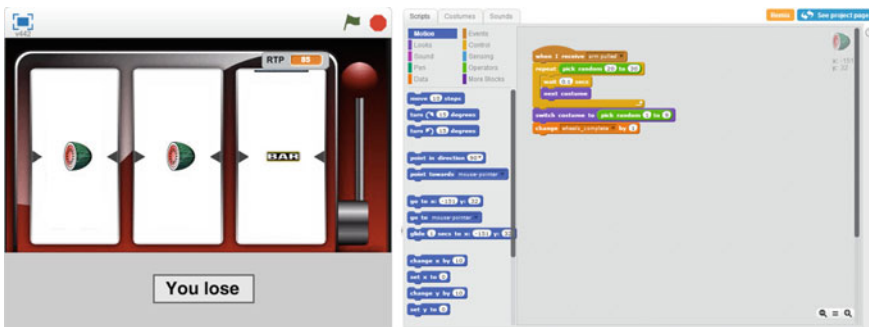


Fig. 16.5 Screenshots from slot machine activity. Students are asked the following question: does the slot machine payback at least 85% of its income?

student needs support in recognizing the nature of the random fluctuation. The best we can do in the MOOC context is to include a self-assessment quiz and offer a review of the lesson. In this example, Noel-Ann shows how she used the app and places some emphasis on the key learning points. Even so, the activities often felt more prescriptive than might have been the case with a teacher working with a class face-to-face, who could respond on the fly to what the students were doing.

The above example is one of many where we adopted the use of technology to instil a sense of playfulness, which we hoped would encourage sustained engagement. Whereas the above activity involves using a specially designed app, we exploited technology in a variety of other ways. For example, an activity on interpreting crime

figures offers the data in the form of a spreadsheet. The data shows the number of violent incidents each year and how many of those were carried out by offenders under the influence of alcohol or drugs. The students are invited to explore this data to investigate possible relationships between crime and the use of alcohol or drugs. We see this example of analysing crime figures as responding to another of the five suggestions by Gal (2002) to work with learners in new ways. The crime figures activity offers an opportunity to develop a critical stance by supporting beliefs with statistical information. In fact, according to the analysis by Watson and Callingham (2003), the use of critical thinking signals the two highest of six levels of statistical literacy (the highest level also requires the use of mathematical expertise such as proportional thinking). A second example is in the playing games PIA of the 'Uncertainty' powerful idea. The students are given a simulation of a slot machine (Fig. 16.5). They are told that slot machines are illegal if they do not pay out in the long term more than a stated proportion of its income, in this case 85%. The activity might support understanding probabilistic aspects of statements about risk and side effects (another of the five new areas in which work might be done with learners, as proposed by Gal 2002). In addition to playing the slot machine game, the student can open up the app and look inside at the coding, written in Scratch (<https://scratch.mit.edu/>).

Being able to open up the app creates new dimensions for an inquisitive student, and builds on earlier stages of Citizen Maths which introduce Scratch programming as a way of revealing hidden mathematics (Noss and Hoyles 1996). Of course, a student who gets out of their depth in this course will have less support to recover than one in a conventional situation. For example, we have noted students becoming frustrated when programming in Scratch because of not knowing simple remedies to problems such as how to clear the screen. Such problems would be resolved trivially by a live tutor. On the other hand, the student can always return to the original app, and try again, although this is not necessarily an ideal solution. In later iterations of the course, we have developed on-screen help and refined its design to improve usability.

There are a number of learning points in this activity. The students may appreciate the power of knowing the underlying probabilities, as can be found by examining code, rather than just running the simulation. They can gain some understanding of how they may win in the short term but they will inevitably lose in the long term if they keep playing. This latter point, as is often the case in Citizen Maths, has social as well as statistical importance.

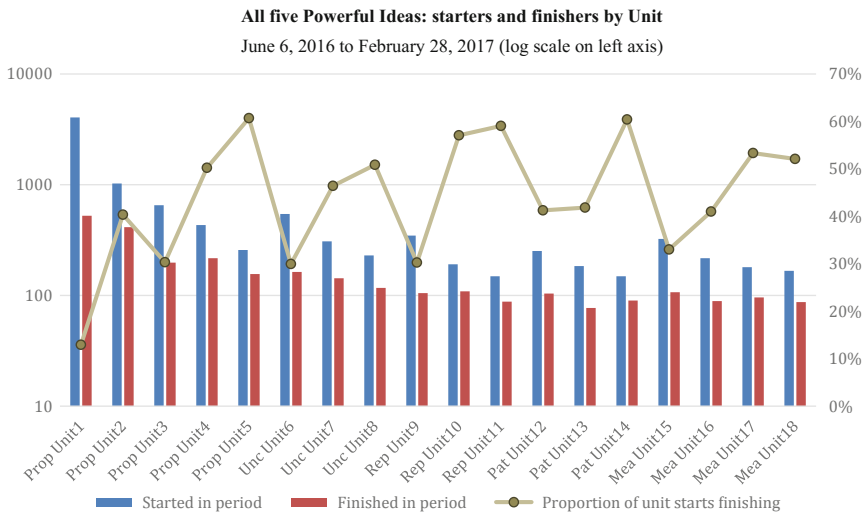
16.3.2 Evaluating Our Response to Challenge 2

We have collected a range of information that we can use to assess engagement. These include the extent to which various part of the course are completed, the views of participants on whether they have met their goals, and feedback from the course 'widgets' on various components of the course.

Figure 16.6 shows a snapshot of course completion during the stage at which all elements of the Citizen Maths course were available. We see that there is drop off on unit completion as the course progresses with relatively small numbers completing all 18 units (less than 100). It is important to note that these data were collected in a 17-week period whereas we expect a typical participant to take as much as 50 weeks to complete the course. As such we expect the unit completion rates to be a little higher as time progresses.

The breakdown of completers by age profile (Fig. 16.7) and gender (Fig. 16.8) compared to registrations is evidence that the course does engage our different groups. The age profile of completers is a little younger than the registrations although all groups have completers. The gender breakdown of completers is similar to that of registrations with a slightly higher proportion of female completers. It is also worth noting that the proportion of completers who identified that they wished to ‘work through the whole course, to improve my maths’ was higher than registrations (85.6% completion compared to 70.1% registrations). This is important as it shows that the course has engaged its target audience more than ‘interested professionals’.

How does the snapshot data (Fig. 16.6) compare to other MOOCs? Clow (2013) notes that progression through MOOCs tends to display a decaying feature, the funnel, far more dramatic than in traditional, face-to-face courses. The author draws together data from a range of sources to illustrate the effect. For example, Clow notes that the “first MIT course, Circuits and Electronics, attracted over 150,000



Prop = proportion, Unc = uncertainty, Rep = representation, Pat = pattern, Mea = measuring (the five Powerful Ideas)

Fig. 16.6 Graph showing a snapshot of the numbers of participants starting and finishing the different elements of the course between June 6, 2016 and February 28, 2017 (note the left hand scale is logarithmic and is used for the starter/finisher bars, the right hand scale is linear and is used to display the proportion of finishers to starters for each unit)

Fig. 16.7 Proportion of registrations and completers by age

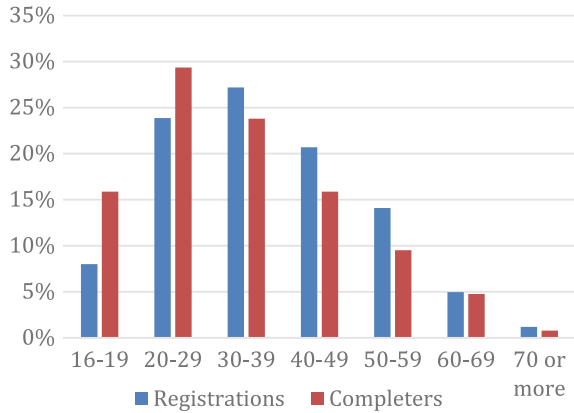
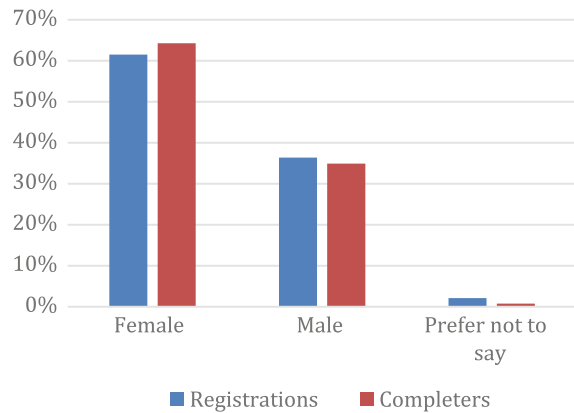


Fig. 16.8 Proportion of registrations and completers by gender



participants, but “fewer than half look at the first problem set”, and only 7157 passed, or about 5%” (Daniel 2012 cited in Clow 2013: 187).

These results suggest that the course is broadly in line with other MOOCs in terms of engagement and completion. Indeed, those who completed each powerful idea were asked to rate the extent to which their goals were met on a scale of 1–5 (1 being “not at all” to 5 “completely”) and were very positive about the course. For the Uncertainty section of the course, the overall figure was just short of 4 and for representation the views averaged 3.5.

At the bottom of each lesson in Citizen Maths is an optional “Rate this lesson” widget, allowing learners rapidly to provide high level feedback as to the utility of the lesson on a five point (0–4) scale (0 being “not at all useful” to 4 “extremely useful”).

Figure 16.9 shows the aggregate feedback for the 18 course units, with the three units within, respectively Uncertainty (6–8) and Representation (9–11) highlighted.

This data shows that learners using the feedback widget are on average reasonably satisfied with the usefulness of Citizen Maths, with some small variations between

successive units. Of course, it must be remembered that widget scores are being provided by a “survivor” population; the further into the course a learner is, the more they might be expected to find course units useful because dissatisfied learners would tend to cease to engage with the course.

At the powerful idea level the data is broken down by age, gender, and goal in Fig. 16.10. Again, we have evidence that across the different groups that we have identified there are participants that have engaged positively with the course.

The completers of each powerful idea were also asked to judge the extent to which they engaged with the course elements. The summary (see Fig. 16.11) shows a broadly positive picture.

There was some differentiation in how each activity type was viewed by participants.

Overall, we see that as the course progresses there is an attrition rate of some concern. Nevertheless, we are optimistic that once the full course has been running for long enough to make a considered judgement, the attrition rate will not be out of line with many other open online courses (see Clow 2013). This is most likely to be true for learners who have made a proper start rather than for those who have merely “dipped in” to take a look. Of those who complete the relevant sections of the course, there is evidence that a good proportion engage well with the various aspects of the course. Of course, we should not be complacent; these positive views are from those who have completed various sections of the course. Those that drop out have not had the chance to contribute to these views. Also, the participants are

Fig. 16.9 Aggregate feedback from the “Rate this lesson” widget

| | |
|-----------------------------|------|
| Unit 1 Mixing | 2.50 |
| Unit 2 Comparing | 2.73 |
| Unit 3 Scaling | 2.46 |
| Unit 4 Sharing | 2.71 |
| Unit 5 Trading off | 2.73 |
| Unit 6 Making decisions | 2.71 |
| Unit 7 Playing | 2.83 |
| Unit 8 Simulating | 2.67 |
| Unit 9 Interpreting data | 2.70 |
| Unit 10 Interpreting charts | 2.80 |
| Unit 11 Comparing groups | 2.84 |
| Unit 12 Appreciating | 2.80 |
| Unit 13 Tiling | 2.76 |
| Unit 14 Constructing | 3.06 |
| Unit 15 Reading scales | 2.86 |
| Unit 16 Converting | 2.73 |
| Unit 17 Estimating | 2.83 |
| Unit 18 Quantifying | 2.97 |

| | Age | | | | | | | |
|-----------------------------|------|-------|-------|-------|-------|-------|-------|------|
| | All | 16-19 | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70+ |
| Uncertainty (Units 6 to 8) | 2.73 | 2.92 | 2.71 | 2.62 | 2.74 | 2.68 | 2.83 | 2.56 |
| Representation (Units 9-11) | 2.77 | 2.99 | 2.84 | 3.04 | 2.57 | 2.70 | 2.66 | 2.28 |

| | Gender | | | Goal* | |
|-----------------------------|--------|--------|------|--------|-----------|
| | All | Female | Male | Goal 1 | Goals 2-6 |
| Uncertainty (Units 6 to 8) | 2.73 | 2.75 | 2.71 | 2.77 | 2.61 |
| Representation (Units 9-11) | 2.77 | 2.82 | 2.64 | 2.96 | 2.32 |

*Goals as in Fig. 16.3. Goal 1 is “to work through the whole course, to improve my maths”

Fig. 16.10 Aggregate feedback from the “Rate this lesson” widget broken down by age, gender, and goal

| Which of the following options best sums up your engagement with ...? | | | | | |
|--|---|---|---|--|---------------------------|
| I carefully watched every video and did every activity, sometimes more than once | I worked through all the content, but sometimes my attention wandered | I skimmed through, mainly to see what was there rather than to engage closely with the course | I sampled what was there, focusing on the things that interested me | None of these options apply in my case | I would prefer not to say |
| Uncertainty | | | | | |
| 36% | 32% | 27% | 5% | 0% | 0% |
| Representation | | | | | |
| 32% | 32% | 27% | 5% | 5% | 0% |

Fig. 16.11 Summary of feedback from users about how engaged they felt by parts of Citizen Maths

not positive across the board and there are some potential concerns that could be addressed. For example, there are some participants who are unconvinced of the value of programming with Scratch.

16.4 Challenge 3: How to Promote Meaningful Mathematics

16.4.1 Discussion of Challenge 3

Many online mathematics learning resources, such as those in the Khan Academy, are closely tied to traditional topic areas—for example, arithmetic, algebra, geometry—and either assume that students will find these meaningful or that other human learning support will provide context and purpose. Citizen Maths aims to communicate the purpose of mathematics in a way that will make it intrinsically meaningful. The challenge to design purposeful activities in a tutor-free environment is further complicated by the need to help the student engage and make sense of the mathematics. In a conventional classroom, effective teachers continuously monitor the student's actions and step in, as necessary, to clarify or offer alternative ways of thinking about the mathematics. Although we have no such opportunity in Citizen Maths, we are dealing with students who, as adult learners, are experienced members of society and will have a range of prior experiences. There is of course some difficulty in exploiting these experiences when the backgrounds of the students will inevitably be so variable. We explored in Citizen Maths solutions in which the technology was adaptive to how the students responded to the challenges. Although such technology is improving rapidly, we were unconvinced that the technical adaptive systems were sufficiently advanced as yet to make effective recommendations to the student in the holistic (as opposed to skill-based) approach we were adopting.

We were aware of the literature on the authenticity trap. Lave's (1988) early work on situated cognition had led to a discussion of the need to create authentic learning experiences; but, it is not possible to take an authentic experience into the classroom, or indeed into a MOOC, because the act of doing so transforms the task, which is no longer authentic. However, research (Nunes et al. 1993) has shown that knowledge is not so much trapped in the situation where that knowledge first became available but is rendered meaningful by that situation. We set out to find situations that were likely to be familiar, even if not directly experienced, and use them to introduce the student to the power of mathematics and statistics in those situations.

The 'interpreting charts' PIA of Citizen Maths has, as one focus, learning about styles and biases in reporting and advertisements, one of the five suggestions for new areas of work with learners proposed by Gal (2002). The following example from this PIA illustrates how it is at the same time possible to engage with charts in personally meaningful ways. This example draws on an applet created by the Institute of Fiscal Studies (www.ifs.org.uk/wheredoyoufitin/). The applet asks the user to enter a few basic facts about themselves, such as their household income, and generates a histogram that shows in which percentile out of the population as a whole the user's net income lies (see the shaded bar in Fig. 16.12 to the left of £300 per week).

In this example, we clearly positioned the learner as the active person at the centre of the task as we imagined they would enter data about their own household. We

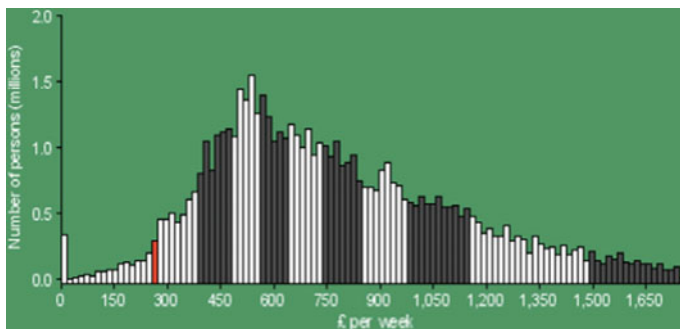


Fig. 16.12 This user's income is well below that of most households in the population

intended that the familiarity of the context would help the student to interpret the histogram. In other situations the learner might feel more like an onlooker. Consider the position of a male student in the following example taken from the 'making decisions' PIA of 'Uncertainty'. The student is given data about the number of women out of 1000 who receive a positive result from a mammogram, used to screen for breast cancer. They are also given data about the proportion of times that the screening machine gives a positive result when the woman does not have breast cancer (a false positive). The result is that, of the women who get a positive result from a mammogram, more than ten times as many do not have breast cancer as the number that do. Although the calculations are not difficult, many people find this result surprising. This situation will be familiar to many students, though may be felt more personally by female than by male students.

This is an example where in a sense we piqued curiosity by courting controversy, a well-known trick that face-to-face teachers use but which we are also able to exploit. The problem is that there is no teacher to support students when the controversy is too upsetting. Might this be problematic for students whose near relative is suffering from breast cancer? Or is there advantage insofar as the student will be less exposed if working with the materials from home on their own rather than in front of peers? We asked such questions in various focus groups. Views were divided but on balance it was felt that the opportunity to raise social awareness should not be missed. After all, it is reasonable to suppose that the use of statistics to support or deny claims during the study of sensitive and controversial topics might promote critical thinking and a broad appreciation of the value of statistics. In the light of these views, we have included such situations in the course, partly because the controversy can be engaging but also because they demonstrate the power of mathematics and statistics to inform such debate. We acknowledge though that this decision might be limiting were the course to be used as part of a blended course or within a learning centre, since controversy raises potentially embarrassing scenarios in collective learning situations that such courses might seek to avoid.

16.4.2 Evaluating Our Response to Challenge 3

One of the key elements of Citizen Maths was the selection of activities that draw upon a range of scenarios in order to engage learners in meaningful mathematics. We noted the example of income data drawn from the Institute of Fiscal studies in ‘interpreting’ and the decision-making aspect of uncertainty using data around mammograms. What evidence do we have to evaluate the extent to which these activities might be perceived as meaningful?

One thing to note is the completion data and positive views of the completers noted above displaying a good engagement by a number of participants. Nevertheless, this does not necessarily mean that the participants feel the mathematics has been meaningful. What we do have are some of the comments made by those using the feedback ‘widgets’. A small number added qualitative comments, many of which were fairly short ‘I did/didn’t find this useful’. Others made some more interesting, and mostly positive, comments.

I really enjoyed this section, I was able to do the calculations quickly & easily. I have even been able to grasp rounding and will be using this method to help me in future tasks.

I feel I have learnt how to read and interpret data on both on a spreadsheet and even without the aid of a spreadsheet. This lesson has been the most enjoyable, clear and easiest to understand.

This is a good app to use to interpret household data. I enjoyed working through the lessons.

And even when the comments suggest some less positive outcomes, there is also evidence of meaningful engagement.

I found this part very difficult to follow but have taken notes so I will be practising this session again.

Of course, there are some more negative comments around some activities. In commenting on a dice simulation activity.

I did not like this activity as it is something I won’t need.

Aside from the comment suggesting that the activity is not meaningful to the learner, going back to challenge 2 and the lack of tutor support, there is no opportunity for us to intervene to discuss this further. A discussion space would offer an opportunity for other users to be involved in debates around such utility issues. Interestingly a more positive comment on the ‘pass the pigs’ game suggests that meaningful activity can be a very personal issue

Definitely got me thinking as I have played the Pass the Pigs so this helped me to construct mental images to accompany the statements about the probability.

16.5 Challenge 4: How to Assess

16.5.1 Discussion of Challenge 4

In designing the structure of the course, we have followed the PISA methodology (www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Mathematics%20Framework%20.pdf) when categorizing activities in terms of content (PISA uses four categories), content topics (15 topics), mathematical processes (3 categories), mathematical capabilities (7 categories), and context (see the four categories above). By using these categories to profile our activities, we have been able to monitor coverage of both mathematical content and the processes without closely prescribing a curriculum, which might have hindered our large grain-size aims.

Citizen Maths does not lead directly to a qualification, although we have collaborated with the awarding organization, OCR, a partner in the project, to track compatibility (or otherwise) between their Cambridge Progression Level 2 units (www.ocr.org.uk/qualifications/by-type/cambridge-progression/) and Citizen Maths. A lack of availability of a direct qualification presumably does not meet some students' desires but our focus has been on students who were motivated to gain a sense of mathematics as a discipline, perhaps as a precursor to working towards a qualification.

Teachers of conventional courses are able to adopt methods of formative assessment to respond to their students within lessons and between lessons in ways not open to us in Citizen Maths. For example, although the course offers a suggested sequence of activities, this preordained sequence might not be suited to some learners because of their prior knowledge or interests. Because there is no tutor to assess whether the proposed sequence should be broken, the course is left open for students to make that decision themselves in the light of progress. Our approach instead was to offer several types of short assessment tasks, aimed at helping the student to make up their own mind whether they have properly understood the content.

These assessment tasks frequently offer multiple-choice questions. When the student gives the wrong answer, responses attempt to give some hint about where they may have gone wrong. When the student gives the correct answer, the response gives a full explanation of the answer in case the student obtained the correct answer through incorrect reasoning. Where possible the wrong options in the multiple-choice questions are chosen to reflect common errors that students make though, in Citizen Maths, the student's wrong responses can only be used for formative assessment by the student and not by a teacher, as recommended by Wiliam (2014).

For example, the activity on interpreting data about incidents of crime, described above, is followed by the question "What is the percentage of alcohol related incidents in 2006/2007?" The correct answer is 54%. It is anticipated that some students will mistakenly refer to the data where the use of drugs (rather than alcohol) is apparent. Such a mistake would lead to an answer of 21%. If a student gives that answer, the following response is given by way of a hint, "This is the percentage of drug related incidents. Use the data for alcohol related".

Multiple-choice questions are not always appropriate to assess learning points. For example, a number of activities in the course have a range of solutions that are not easily captured by a system of offering multiple choices. In other cases, the essence of the task lies in the process of doing it, rather than in the outcome. It is not perhaps surprising that this type of activity is quite prevalent when the aim of the course is to engage students in work that reveals the power of mathematics. In such cases, the onus is on the student to watch the to-camera tutor review the task and decide for themselves whether they have understood it sufficiently. One of the powers that Citizen Maths and other MOOCs have that is not available in the conventional classroom is that students can easily return as many times as they want to a video of the tutor's explanation or to a task they have already done. So, we hope that a student who is not satisfied that they have properly understood the activity after watching the tutor is able to return to the introductory video and the task itself.

16.5.2 Evaluating Our Response to Challenge 4

We have no direct evidence for the extent to which these elements of the course have been effective. The multiple-choice questions are typical activities within lessons, not a separate element. During the development of the course, we chose to collect data on the lesson level rather than on individual components. As such we cannot claim any particular success for the responses to participant choices except to note that the sessions overall have achieved good ratings, the quizzes are noted as a good feature (see Fig. 16.13) and there are positive comments from users about the value of the quizzes. One learner describing the positive features of Citizen Maths noted that with the quizzes there is a “summary of the question you’ve just answered and whether you got it correct, and if you haven’t, an explanation of why it’s not correct”.

16.6 Final Comments

We accept that some of our solutions have been compromises in the sense that they often do not match our pedagogic ideals, which we have learned through research and teaching in conventional classrooms. Some colleges of further education are using our materials as additional resources and in those cases a tutor will typically be able to support the learning process. On the other hand, it is not realistic to expect the shortfall in Level 2 mathematical understanding across the population to be remedied in conventional ways. The 2011 Skills for Life survey (BIS 2012) estimated that some 72% of the population of England were below Level 2 and the 2012 international survey of adult skills (PIACC) (BIS 2013), while using different level descriptors, was consistent with this level of need. The existing educational structure of further, adult and vocational education could not cater for anything close to numbers on this scale. This problem may be even more acute in the case of statistical literacy,

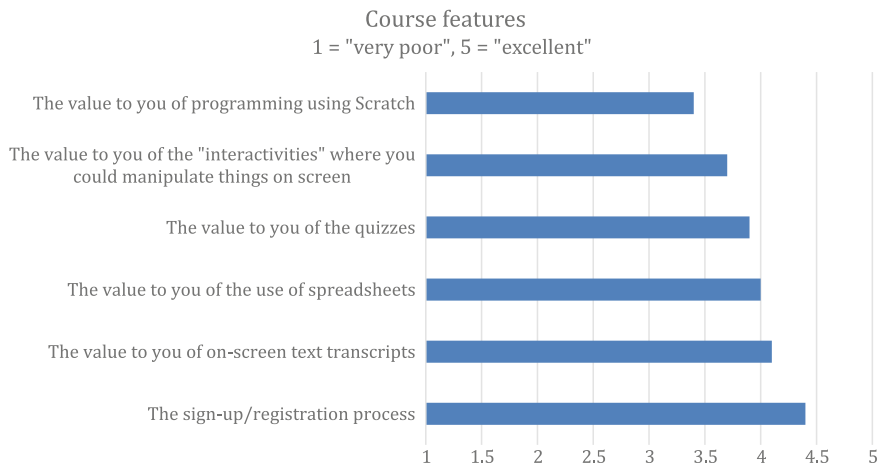


Fig. 16.13 Feedback on Citizen Maths course features by users

where the shortage of appropriately knowledgeable teachers is widely reported (see, for example in [Batanero et al. 2011](#)) and where a better-informed populace would advantage individuals and society (see, for example, [Gal 2002](#)).

Designing the structure and the content of Citizen Maths has been an interesting challenge and we have described some of those challenges in this paper. By focussing on powerful ideas rather than specific small grain-size techniques, we have placed emphasis on a broad interpretation of numeracy and statistical literacy. While it could be argued that such a change in emphasis would benefit teaching and learning in conventional classrooms, it seems critically important in a MOOC-based approach where the students are heterogeneous and sufficiently mature to bring rich experience to the learning environment.

We believe that the use of contextualised activities actualises those powerful ideas in familiar and meaningful ways whether those tasks are fundamentally mathematical or statistical in nature. Of course, in the representation and uncertainty PIAs, the role of context potentially extends beyond motivation and meaningfulness to opportunities for developing understanding of the contextually-dependent transitions from design to data capture, from exploring data to identifying patterns and towards the end of the statistical problem-solving activity from results to their interpretation ([Cobb and Moore 1997](#)). We note though that the transition from design to data would be challenging to implement in a MOOC and has not as yet been the focus for an activity in Citizen Maths.

We have experienced affordances in the use of a MOOC that are common with the use of technology in teaching and learning statistics and probability in classroom contexts: (i) the facility to handle complex computations can help the student to focus on the underlying conceptual challenge and not be distracted by the difficulties of carrying out calculations; (ii) the repetition of trials digitally enables data to be collected quickly so that variation in data can be explored; (iii) the use of computer-

based simulations allows the learner to explore situations which might not otherwise be accessible to the student; (iv) programming the computer allows the learner to try out models of solutions and then modify those proposed solutions according to feedback; (v) dynamic graphical displays tends to support visualisation of mathematical and statistical concepts.

Of course, we do not argue that MOOCs such as Citizen Maths offer as good a learning experience as can be provided by teachers, but we have noticed some advantages: (i) students can work intensively over a short spaces of time or spread their work over longer periods to fit learning around other commitments; (ii) they can learn at their own pace by repeating lessons or specific videos as often as they like; (iii) students can make choices about what content to study and in which order, which might suit those students who are aware of particular strengths and weaknesses in mathematics or statistics; (iv) it is possible to access the resources remotely rather than needing to travel to a particular location to engage in learning; (v) some students might appreciate being able to learn on their own in order to avoid embarrassment should they feel inadequate in some way or if the topic were in some way sensitive or controversial (though of course there are also distinct advantages to collaborative learning, which is more natural in a conventional classroom).

As social platforms continue to improve it is possible that MOOCs will in the future be able to support collaborative learning in ways that have not been built into Citizen Maths. Some might argue that future research might provide a better understanding of conceptual development in mathematics and statistics with opportunities for artificially intelligent support structures. Our experiences suggest however that when the focus is on broad interpretations of numeracy and statistical literacy, the grain size of what needs to be learned is large enough for us to believe that society is still some way from being able to develop such systems.

The data collected so far on users offers some broadly positive feedback about the extent to which we met these challenges. Sustaining Citizen Maths in the future will be another challenge. Some additional finance will be needed as shortcomings in the current course will inevitably be identified and the use of data files will quickly need to be updated. Fundamental changes in technology will offer new solutions, including better ways to personalize the experience, making old technologies look obsolete. Citizen Maths was never going to be an ideal solution but the process has enhanced our knowledge about what seems to work and what does not, knowledge which should help us be ready for that time.

16.7 Endnote

This paper is narrowly focused on the interests of Topic Study Group 15. But a wider range of issues appears when creating (maths) MOOCs. Several of these are described on the Citizen Maths web site, at <https://citizenmaths.com/>, and, in particular, on the Citizen Maths blog, at <https://citizenmaths.com/blog>.

Acknowledgements We also wish to acknowledge the Ufi Charitable Trust who funded the work to develop the Citizen Maths course.

References

- Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal*, 32(1), 23–38.
- Batanero, C., Burrill, G., & Reading, C. (Eds.) (2011). *Teaching statistics in school mathematics—Challenges for teaching and teacher education: A joint ICMI/IASE study: The 18th ICMI study*. <http://link.springer.com/book/10.1007%2F978-94-007-1131-0#section=933243&page=1&locus=0>. Accessed 14 January, 2016.
- BIS. (2012). *2011 Skills for life survey: A survey of literacy, numeracy and ICT levels in England*. BIS Research Paper Number 81. London: Department for Business Innovation and Skills.
- BIS. (2013). *The international survey of adult skills 2012: Adult literacy, numeracy and problem solving skills in England*. BIS Research Paper 139A. London: Department for Business Innovation and Skills.
- Bosworth, D., & Kersley, H. (November, 2015). *Opportunities and outcomes in education and work: Gender effects, Research briefing*. UK Commission for Employment and Skills. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/477360/UKCES_Gender_Effects.pdf. Accessed 26 October, 2016.
- Clow, D. (2013). MOOCs and the funnel of participation. In D. Suthers, K. Verbert, E. Duval & X. Ochoa (Eds.), *Proceedings of the Third Conference on Learning Analytics and Knowledge (LAK 2013)*, 8–12 April 2013, Leuven, Belgium (pp. 185–189).
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104(9), 801–823.
- Daniel, J. (2012). *Making sense of MOOCs: Musings in a maze of myth, paradox and possibility*. <http://sirjohn.ca/wp-content/uploads/2012/08/MOOCs-Best.pdf>. Accessed 29 May, 2017.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1–25.
- Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. In M. Sahami (Ed.), *Proceedings of the first ACM Conference on Learning @ Scale Conference*, 4–5 March, Atlanta, GA, USA (pp. 41–50).
- Lave, J. (1988). *Cognition in practice*. Cambridge: Cambridge University Press.
- Noss, R., & Hoyles, C. (1996). *Windows on mathematical meanings*. Berlin: Springer.
- Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics*. Cambridge: Cambridge University Press.
- Papert, S. (1972). Teaching children to be mathematicians versus teaching about mathematics. *International Journal of Mathematics Education and Science and Technology*, 3, 249–262.
- Pratt, D. (2012). *Making mathematics phenomenal*, based on an Inaugural Professorial Lecture delivered at the Institute of Education, University of London, on 14 March 2012. Institute of Education, University of London, Professorial Lecture Series.
- Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, 57(2), 74–79.
- Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1–8.
- Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3–46.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–248.
- William, D. (2014). The right questions, the right way. *Educational Leadership*, 71(6), 16–19.

Chapter 17

Critical Citizenship in Colombian Statistics Textbooks



Lucía Zapata-Cardona and Luis Miguel Marrugo Escobar

Abstract The goal of this research is to study how the statistical component of fifth-grade mathematics textbooks in Colombia contributes to the development of students' critical citizenship. This research followed a socio-critical perspective. Content analysis was the technique used to make sense of the data produced by analyzing seven mathematics textbooks, and the units of analysis were 261 tasks in the statistical components. The results show that the contexts of the tasks were mostly hypothetical with very few tasks presented in real contexts. The tasks mainly functioned as platforms to introduce measurement calculations and application of statistical procedures. When the tasks were presented within real contexts, the conditions were not used to the fullest extent to contribute to the development of students' critical citizenship. The tasks promoted mainly statistical and technological knowledge over reflective knowledge, failing to contribute to the students' socio political awareness.

Keywords Critical citizenship · Statistics education · Textbooks · Textbook tasks

17.1 Introduction

Literature suggests that students in compulsory education learn a great deal of information about statistical concepts and procedures that they are unable to use when confronted with real world problems (Bakker et al. 2012). Students learn to manipulate statistical symbols to perform well on school exams, but such practices are useless in out-of-school situations (Roth 1996). This phenomenon reveals a dichotomy between school knowledge and the student world, which is promoted mainly by the way in which statistics is taught—fragmented and procedural [some authors call it inert knowledge (Bakker and Derry 2011)] and disconnected from the context in

L. Zapata-Cardona (✉) · L. M. Marrugo Escobar
Universidad de Antioquia, Medellín, Colombia
e-mail: lucia.zapata1@udea.edu.co

L. M. Marrugo Escobar
e-mail: luis.marrugo@udea.edu.co

© Springer Nature Switzerland AG 2019
G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_17

which knowledge is produced. In an era of information and technology, schooling does not need to focus on factual knowledge; instead, it could focus on developing students' critical citizenship—a quality of thought that supports students “to be critical citizens who can challenge and believe that their actions will make differences in society” (Giroux 1988, as paraphrased in Skovsmose 1992, p. 2). To accomplish this goal, many authors have argued that teaching should encourage the mathematical process of modelling in real contexts (Barbosa 2006; Biembengut and Hein 2002; Skovsmose 1992; Stillman et al. 2013) imbedded in critical issues of society (in the sense proposed by Skovsmose 1999), and we support that argument. Critical citizenship emphasizes awareness of the social/political context (Stillman et al. 2013), promotes environmentally and socially aware citizens, and develops a critical disposition towards the surrounding world.

Textbooks are essential tools in the educational system in different societies. Teachers and students as well as parents use textbooks as curricular guidelines. For many teachers, textbooks provide resources for the design, implementation and assessment of statistics lessons and can influence the way teaching practice is oriented. Particularly in Colombia, teachers who have the responsibility of teaching statistics are mathematics teachers with a very unsteady statistics background, and textbooks can give them the security they lack. Taking into account the role of the textbooks as essential materials in the educational system, we explore the relationship among the statistical components of mathematics textbooks and students' opportunities to develop socio-political awareness. The research question explored in this research is: To what extent does the statistical component of fifth-grade mathematics textbooks in Colombia have the potential to play a transformative role in developing students' critical citizenship?

Colombian curriculum guidelines do not specifically address the development of critical citizenship. In fact, Colombian curriculum guidelines are more focused on making explicit the minimum performance standards and contents that students have to master at the end of the school year than on developing critical thinking as members of society. However, positioned in a critical epistemology of knowledge, we consider that, “[...] education must take part in efforts to educate students to be critical engaged citizens” (Barbosa 2006, p. 294). In other words, the development of critical citizenship is the ultimate goal of education. Students' knowledge should not be isolated from the social and cultural environment in which they live. There is a very shortsighted view when the curriculum focuses exclusively on procedures. Nowadays, statistics education in many places in the world is focusing attention beyond being a simple methodological science. For example, the mission statement of the American Statistical Association includes the goal of “using our discipline to enhance human welfare” (Lesser 2007). This statement means that statistics should be a tool to help citizens understand and transform their world in ways that go beyond just learning statistical concepts and procedures. The goal of the work described in this chapter is to make public the opportunities to develop critical citizenship from curriculum materials in Colombia.

17.2 Theoretical Framework

17.2.1 Critical Citizenship

We understand critical citizenship as an intellectual tool oriented to educate critical and aware citizens who have the responsibility to participate in society and contribute to its transformation. Critical citizens go beyond mastering statistical (in general, scientific) skills and use them to understand and transform critical issues in society. Some authors have referred to the same construct but using different language. For example, Giroux (1988) talks about *critical democracy* to refer to the possibilities of thought—rooted in a spirit of critique—that enables people to participate in the understanding and transformation of their society. D’Ambrosio (1999) and Skovsmose (1992)—possibly inspired by the ideas of *literacy* from Paulo Freire (Guerrero 2008)—talk about *mathemacy* and *matheracy* respectively. In their discussions, *mathemacy* is the mathematics for equity and democracy. It is a fusion of mathematics with democracy to educate students to be critical, engaged citizens. *Matheracy* offers “a much broader dimension to mathematical thinking, stressing its value as an instrument for explaining, understanding and coping with reality in its broad sense. Matheracy is the main intellectual instrument for the critical view of the world” (D’Ambrosio 1999, p. 150). *Matheracy* has a strong connection with the real context—called the “world outside” by D’Ambrosio.

Barbosa (2006) assumed a *socio-critical perspective* in which the teaching of mathematics should contribute to the critical understanding of the surrounding world and should promote reflection on the role of mathematics in society. From this perspective, one of the aims of mathematics in school is to produce critical, politically engaged citizens. According to Barbosa, the context needs to be extracted from students’ everyday lives or other sciences that are not pure mathematics. It is essential that the contexts are real instead of hypothetical or fictional. The term *real* within a socio-critical perspective means that the context is tied to social facts and critical issues of society. The real context has to touch students’ lives because the meanings are negotiated, and the inferences are linked to the students’ specific real-world problems. Stillman et al. (2013) use the expression *socio-critical competency* to refer to the development of critical dispositions emphasizing awareness of the social/political context and state that the contexts related to social, economic or environmental indicators have potential for enhancing an understanding of the world.

Skovsmose (1992) refers to a similar construct called *democratic competence—critical citizenship*—when discussing the development of the required skills for “children’s and adolescents’ later participation in democratic life as critical citizens” (p. 3). Finally, Valero (2002) calls it *citizenship education*¹ referring to the development of citizen awareness. She argues that citizens do not act in the world exclusively in terms of their cognitive dimension, but they participate in a social-

¹This is a translation by the authors. The expression in Spanish is *Formación ciudadana*.

economic-politic-historic-cultural world. Her point of view implies the inclusion of real references related to social facts or problems in the teaching of mathematics.

This literature review shows that critical democracy, matheracy (or mathemacy), socio-critical competency, critical citizenship, and citizenship education are different expressions referring to a similar notion, that in this paper is called critical citizenship. All the authors agree with the importance of connecting the school to the “world outside” through real contexts. According to Skovsmose (1992), in the process of developing democratic competency (critical citizenship), three types of knowledge come into play: *mathematical knowledge*, *technological knowledge* and *reflective knowledge*. To develop critical citizenship, reflective knowledge should be privileged over mathematical and technological knowledge. Mathematical knowledge is linked to the mathematical skills, including competencies to reproduce mathematical thinking, theorems, and proofs as well as mastering a variety of algorithms. In this research, the *mathematical knowledge* is the statistical knowledge (i.e. required knowledge to calculate a central tendency measurement). *Technological knowledge*² is the knowledge needed to develop and use the technology (i.e. required knowledge to represent data in a graphic). *Reflective knowledge* is meta-knowledge based on broad interpretations and previous knowledge. The starting point is to understand the situation in which the technological knowledge comes into play, but the goal is the reflection on the ethical and social consequences of technology in society (i.e. assessment of the social consequences of a study that uses per capita income of a population as an estimate of the quality of life when there are strong inequities within the members of the community). *Reflective knowledge* could lead to different kinds of questions: Are we using the appropriate statistical summary? Are the results reliable? Can we use the result effectively? Can we say something about the quality of life using other methods? *Reflective knowledge* takes into account the mathematical and the technological knowledge.

17.2.2 *Context in Developing Critical Citizenship*

The statistics education community has conventionally valued the context of the data in teaching statistics. Cobb and Moore’s statement about the important of context in statistics is well known: “data are not just numbers, they are numbers with a context!” (1997, p. 801). In that regard, according to Pfannkuch and colleagues (2016), the modeler goes back and forth between mathematics and reality in the modelling activity, and the context is the source of meaning (Cobb and Moore 1997). However,

²Skovsmose uses the context of motoring to clarify the differences among technological and reflective knowledge. Technological knowledge is related to the skills required to construct and repair a car while the reflective knowledge is related to the skills required to drive a car and to assess the social consequences of car production. “Technological knowledge itself is unable to predict and analyse the results of its own production; reflections are needed. The competence in constructing cars is not sufficient for the evaluation of the social consequences of car production” (Skovsmose 1992, p.7).

in a socio-critical perspective, the context goes beyond the scope that the statistics education community traditionally has given to it.

In the socio-critical perspective, the development of critical citizenship has a close relation to the context in which the tasks are imbedded. According to Stillman et al. (2013), real contexts are fundamental to promote the development of critical citizenship. Other authors have stated that framing statistics in real world situations could constitute a way to study and transform critical issues in society. Thus for example, Lesser states “some datasets from the real world may have the power to effect a lasting appreciation of or even commitment to statistics as a tool to help understand (and maybe improve) some of our society’s most profound or pressing matters” (2007, paragraph 1). The context is not only an excuse to frame statistical problems in attractive ways but a setting in which students learn about the world in which they live, explore it empirically and get tools to critically act and react. While this process of developing awareness is taking place, students learn content and statistical tools to make sense of their world. Thus, *reflective knowledge* is the ultimate goal in the socio-critical perspective while *statistical* and *technological knowledge* are simultaneously developed.

The purpose of curriculum materials is to support teaching, facilitate students learning (Travé González and Pozuelos Estrada 2008) and help students to master skills (Mateos Montero 2008). When an additional goal to support students transforming critical issues of society is added, context becomes essential. Curriculum materials based on social, economic and environmental contexts allow students to develop skills in the management of statistical tools but also to form their social consciousness. In this sense, the formative power of statistics and its connection to human activity is emphasized. Statistics is the result of human construction; therefore, it is an activity that responds to the requirements of culture and is born of the needs of the human being. “The objects and the scientific activity are social and cultural results” (Etchegaray 2010, p. 14); therefore, curriculum materials to support the teaching of statistics should not be isolated from the human condition, nor from social and cultural contexts. In general, sciences are a response to problems of anthropological nature that arise in relation to the physical, biological and social environment in which the human being lives (Etchegaray 2010). Taking into account this consideration, curriculum materials to teach statistics should address this relationship of the human being with nature (in the sense expressed by Radford 2016).

Accumulation of statistical knowledge is not enough for developing critical citizenship. To accomplish this goal, tasks should be framed in contexts that students can study critically. Curriculum materials should include nontrivial tasks related to critical issues of society (environmental problems, social inequalities, gender bias, social indicators) offering opportunities for students to reflect upon the context of these tasks as they learn or apply the associated statistical content, procedures and tools. Teaching statistics using critical issues of society “would also include students developing a sense of empowerment to be able to use statistics to “talk back” to or change the world” (Lesser 2007, paragraph 9).

Framing tasks within real contexts could function as a strategy to support the development of students’ *reflective knowledge*. Such a strategy allows students to

move back and forth between the real world and the formalities of the statistical science. This strategy can also help to overcome inert, fragmented and disconnected statistical knowledge—disengaged from the student world—and contribute to the understanding and transformation of society. It has the potential to link concepts, statistical reasoning, investigative processes, and real social contexts.

17.2.3 Context in Developing Statistical Reasoning and Thinking

Framing tasks in real contexts to teach statistics is not only important for the development of critical citizenship but also for the development of statistical thinking and statistical reasoning. Literature has shown that traditional approaches to teaching statistics have focused predominantly on skills, concepts and procedures, failing to lead students to think and reason statistically (Ben-Zvi and Garfield 2004). Using problems in real contexts to teach statistics could have the potential to promote statistical thinking and reasoning.

In this regard, Pfannkuch and Wild (2004) have indicated that one approach to contribute to students' development of statistical thinking is by solving real world problems. They suggested that real world problems help students in making predictions, seeking explanations, finding causes within concrete contexts. According to the authors, statistical thinking is a construct related to understanding the big ideas about variation, sampling, and transnumeration, but it also includes being able to understand and utilize the context of a problem in forming investigations and drawing conclusions. Pfannkuch and Wild talk about statistical thinking as an integration of contextual knowledge and statistical knowledge. The contextual situation permeates the statistical summaries and any type of analysis. Statistical thinkers are able to make inferences from data but also to critique and evaluate the result of a problem solved.

According to Garfield (2002), and later expanded by Ben-Zvi and Garfield (2004), statistical reasoning may be defined as the way people reason with statistical ideas and make sense of statistical information. This involves making interpretations based on sets of data, representations of data, or statistical summaries of data. Reasoning means understanding and being able to explain statistical processes and being able to fully interpret statistical results within a particular context. Both statistical reasoning and thinking have strong connections to the context.

17.2.4 Research in Textbooks

A textbook is an extensive printed object intended to guide the student's work throughout the school year in a specific area of knowledge (Johansson 2003).

Research in mathematics textbooks is a field with a recent history, and much of the development of this line of research has taken place in the last three decades (Fan et al. 2013) and has moved in different directions. Some studies have explored the organization of the content presented in the textbooks (González-Astudillo and Sierra-Vasquez 2004; Jones and Fujita 2013; Mesa 2000; Otte 1986; Salcedo 2015). Other studies have set their sights on the coherence among textbooks and public policy and research (Borba and Selva 2013; Cantoral 1997; Chandler and Brosnan 1995; Johansson 2003; Usiskin 2013); others have studied how students (Rezat 2013) and teachers use the textbook (Nie et al. 2013; Remillard 2005). Other studies have focused on historical studies of textbooks (González-Astudillo and Sierra-Vasquez 2004; Howson 2013; Pepin et al. 2013; Xu 2013). Finally, another group of studies examined textbooks from critical standpoints or in other words, the possibilities of the textbooks to contribute to the development of the social dimension of the subjects who use them (Herbel-Eisenmann 2007; Stillman et al. 2013; Österholm and Bergqvist 2013). This suggests that while research about textbooks has focused on specific aspects of the texts such as content, use, coherence with public policy, or historical revision, very few studies have conceived the textbook as an instrument with transformative capacity in the educational process. Some authors have considered it “excessive to believe that a textbook can or should even cause a vital transformation. It is one of the materials provided to use in the classroom, but its influence should not be overestimated”³ (Prendes-Espinosa 2001). Contrary to this approach, we feel that the textbook is not only an instrument to disseminate objective knowledge—technical and intransformable—that students have to assimilate, but it should be a tool to support learning as a “communal acquisition of forms of reflection of the world” (see Footnote 3) (Radford 2006, p. 114), and “the process that constitutes our human capacities” (p. 114). Consequently, it is necessary to deepen the study of the textbook as a tool that supports the development of students’ critical citizenship. This research is a contribution in that direction.

17.3 Methodology

The research question explored in this study is: to what extent does the statistical component of fifth-grade mathematics textbooks in Colombia have the potential to play a transformative role in developing students’ critical citizenship. To answer that question, it is necessary to study what is in the textbooks and to study the context in which the tasks are proposed. For that reason, the technique to make sense of the data collected was content analysis. Although content analysis can be systematic, objective and quantitative, it moves between “the rigor of objectivity and the fertility of subjectivity” (López-Noguero 2002, p. 173), and attempts to uncover the hidden, the “unsaid” of the documents. It reveals the internal structure of the information on the sources of study.

³Translation from the authors.

The documents studied were the statistical components of fifth-grade (age 10 and 11) mathematics textbooks available in printed format in the textbook market of the Colombian educational system. Fifth grade is an intermediate moment in Colombian schooling (10–11 years of age) that is appropriate to study textbook contribution to the development of critical citizenship. The very first grades in the Colombian educational system are dedicated to students' integration into the school life, and although there is use of textbooks, the use is sporadic and occurs under the close orientation of teachers. Literature has shown that this is a common phenomenon, and that from 4th and 5th grades school mathematics—both content and organization—is essentially based on the use of the textbook (Johansson 2003).

Even though we recognize that there are different formats of textbooks and that the electronic texts are gaining recognition, we focused on the most popular format in our socio-cultural context, which is the printed version. We selected the textbooks using the textbook catalogue⁴ prepared by the Ministry of Education of Colombia to help teachers and schools selecting their textbooks. The catalogue listed 28 mathematics textbooks from ten different publishers from 2005 to 2014. If a publisher had several editions of a textbook in that 10-year period, we only took into account the most recent edition. Although the catalogue included textbooks in English for bilingual education, we only considered textbooks in Spanish since it is the official language of instruction. The catalogue listed some textbooks that were not available in the market any longer. We only included those available in the textbook market at the beginning of 2014. We validated the information with two bookstore visits.⁵ Table 17.1 presents the selected textbooks.

The units of analysis were the tasks in the statistical components of the mathematics textbooks proposed for the students. A statistical task is what the student was asked to do, such as: application of algorithms, symbol manipulation, design representations, transformation of problems in expressions or models (Shimizu et al. 2010). In other words, it is the assignment proposed for the students in the textbook after introducing a statistical topic. We were not interested in how the tasks were enacted in class but in studying the tasks themselves, as they seem to be intended by the designers, in terms of the potential role in developing students' critical citizenship. The use of the task as the unit of analysis is justified according to Trends in International Mathematics and Science Study—TIMSS 1999 in which more than 80% of the time of the regular mathematics class students spend it in mathematical tasks (Hiebert et al. 2003). An example of a task appears in Fig. 17.1. We designed and applied an instrument to each of the 261 tasks proposed in the statistical component of the selected textbooks. We did not use any criterion either to include or exclude tasks; every task was analyzed. The instrument focused on the *nature of the context* (real, hypothetical, theoretical, without context), *dimension of context*

⁴Catalogue was available on line, at the beginning of 2014, in the following link http://aplicaciones2.colombiaaprende.edu.co/textos_escolares/contenidos/resultado_busqueda.php. However, the link is not available anymore.

⁵One of the bookstores has its catalogue available on-line. Catalogue visited on February 11, 2014 <http://www.panamericana.com.co/categorias/categoria.aspx?p=reN6vDa2UjMBxvdrQqQYWeLel8MiV6lZ>

Table 17.1 Selected textbooks

| Textbook name [Translation into English] | Publisher |
|--|---------------------|
| Los Caminos del Saber 5 (Joya-Vega, et al. 2014) [The Paths of Knowledge 5] | Santillana |
| Avanza Matemáticas 5 (Silva-Calderon, 2014) [Forward Mathematics 5] | Norma |
| Enlace Matemáticas 5 (Acevedo-Cacedo and Pérez-de-Díaz, 2013) [Mathematics Link 5] | Educar |
| Matemáticas en Red 5 (Duran et al. 2013) [Mathematics in the Net 5] | Ediciones SM |
| Matemática Experimental 5 (Uribe-Cálad, 2013) [Experimental Mathematics 5] | Uros Editores |
| Zona Activa 5 (Mejía-Fonseca et al. 2011) [Active Zone 5] | Voluntad |
| Contacto Matemático 5 (Beltrán-Beltrán et al. 2014) [Mathematics Contact 5] | Editorial Educativa |

(political, social, economic, scientific, historical, cultural), *type of knowledge promoted* according the classification of Skovsmose (1992)—statistical, technological, or reflective. As we mentioned before, *reflective knowledge* could combine *statistical* and *technological knowledge*. Consequently, a task could have been doubled coded; however, we only coded the most prevalent *type of knowledge*. A task in a *real* context gives accurate information that can be contrasted with external sources—data bases from statistical departments, historical records, or scientific data, for example: “number of kilometers different vehicles travel with a gallon of gasoline” (Acevedo-Cacedo and Pérez-de-Díaz 2013, p. 67). A task in a hypothetical context provides no verifiable information, for example: “the age of young people registered for a contest was recorded and presented in the diagram. Construct the frequency table” (Beltrán-Beltrán et al. 2014, p. 257). The data of a task in a theoretical context can be simulated by using manipulatives or software or by applying combinatorics theory; for example, “write all the pairs after throwing two dice” (Joya-Vega et al. 2014, p. 244). A task without context is presented away from any connection to the world; for example, “write the mode of each data set, $A = [2, 3, 4, 4, 5, 6, 7]$ ” (Joya-Vega et al. 2014, p. 180). The *dimension of the context* was only analyzed in the tasks of *real nature*. The *dimension* has to do with the range over which the real tasks extend (social, political and others). As a validation strategy, the two researchers applied the instrument to each task independently and then compared the coding. The results revealed a high level of agreement in the coding; those in which the researchers disagreed were resolved by discussion. To illustrate the application of the instrument, Fig. 17.1 shows an example of a task. In that example, the *nature of the context* of the task was coded as *real*, since the exchange rate from US dollar (USD) to Colombian peso (COP) can be contrasted with historical reports in newspapers. The *dimension of the context* was coded as *economic*, since the task relates two well-known currencies.

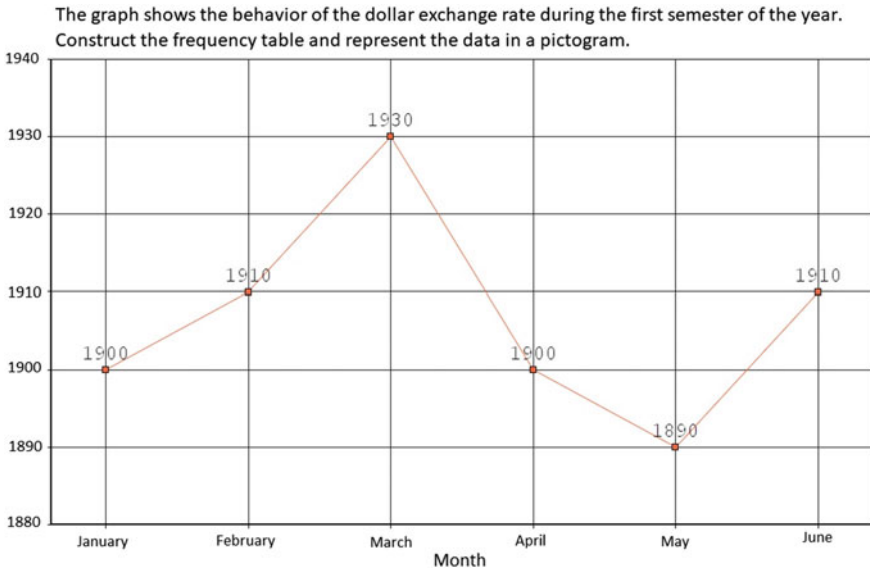


Fig. 17.1 Example of a task taken from Beltrán-Beltrán et al. (2014, p. 254)

The *type of knowledge* promoted was coded as *technological* (knowledge needed to develop and use technology) since the task asked the students to use information already given to construct a new graphical representation.

17.4 Results

By studying the *nature of the context* linked to the tasks (Fig. 17.2), we found that the majority were hypothetical (59.3% of the total, 155 tasks), and very few were real (11.5% of the total, 30 tasks). This finding was similar when we analyzed the nature of context discriminated by textbook (Fig. 17.3). Although the textbooks produced by Norma Publisher (Silva-Calderon 2014) and by Voluntad Publisher (Mejía-Fonseca et al. 2011) stood out among the rest with a larger number of tasks in real contexts, the number of these was still small compared to the tasks in hypothetical contexts.

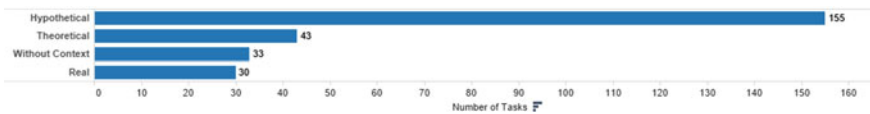


Fig. 17.2 Nature of the context of the tasks in textbooks



Fig. 17.3 Nature of the context of the tasks by Publisher

Taking into account the socio-critical perspective, this is a troubling result since real contexts are essential to promote the development of critical citizenship. We cannot understand the world by turning our backs to the realities of the world. The problem with non-real (hypothetical and theoretical) or absent contexts is that instead of helping students improve their understanding of the world, such contexts offer a strange view of the reality and a limited perspective of the implementation of school mathematics in the world (Stillman et al. 2013). Furthermore, the tasks in non-real contexts keep promoting the division between the world and the school, and students continue to believe that school knowledge is only useful and functional in school and that the world needs another type of knowledge. To develop a critical view of the world students should be given tasks that model social phenomena, where context is fundamental.

Although the tasks presented in real contexts were unusual, those were analyzed according to the *dimension of context* (see Fig. 17.4). Most of the tasks in real contexts were linked to sports (e.g., number of athletes participating in the National Games in 2012, by gender, by state and by sport category; medals won at the National Games in 2012 by state; medals won by athletes at the Pan American Games in Rio de Janeiro in 2007) followed by scientific contexts (e.g., life span of animals, fuel consumption by type of vehicle, number of species per fungi group, weight of an elephant by numbers of days after being born), social contexts (e.g., production of waste by country, number of people per state in Colombia) and artistic contexts (e.g., number of strings by instrument in the National Symphony Orchestra, average time to repeat a tongue-twister poem in the classroom, number of magicians attending the International Day of Magic in Spain). Very few tasks were presented in economic contexts (e.g., variation of the price for a pound of coffee from 2004 to 2009, minimum wage in Colombia from 2005 to 2009, performance of the US dollar exchange rate in the first half of the year), and political contexts. Even though some of the real contexts were interesting, the context was exclusively used to introduce statistical measurement calculations or the construction of graphical representations. There was not a single reference in which tasks in real contexts were used as an avenue to enhance students' critical citizenship.

To contribute to the development of the students' critical citizenship, understanding and involvement in the world, the tasks should be presented in real contexts that reflect critical issues of society. Tasks related to population growth, food production, spread of diseases, climate change, diet, quality of life, poverty, poverty measures, socio-economic indicators, human development indicators, environmental impact have the potential to contribute to these objectives (Stillman et al. 2013).

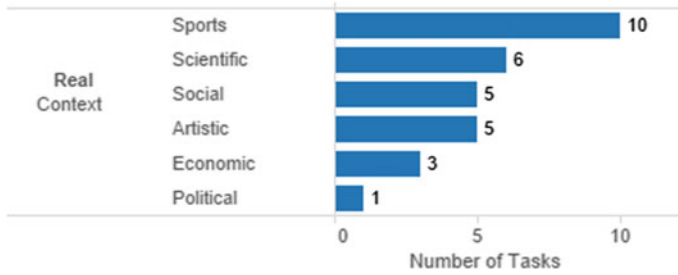


Fig. 17.4 Dimension of real contexts within the textbooks

Another interesting result was the *type of knowledge* promoted for the tasks. A task can promote statistical, technological or reflective knowledge (or more than one simultaneously). The socio-critical perspective focuses primarily on promoting reflective knowledge where the statistical or technological knowledge are associated with reflective knowledge. We found that the textbooks tasks promoted essentially statistical knowledge (60.9% of the tasks), followed by technological knowledge (37.9% of the tasks). Only three tasks (1.1%) were focused on stimulating reflective knowledge (Fig. 17.5). The development of critical citizenship is associated with the opportunities students have to engage in reflection; these opportunities are diminished if there is little chance to connect students with their world.

The following is an example of a task in one of the textbooks that promoted reflective knowledge: “Identify a problem in your school. Carry out a statistical process. Use the analysis of the information to generate strategies to solve the identified problem” (Duran et al. 2013, p. 51). The task is very simple, but it articulates several elements (1) positions the task in a real context that is familiar and pertinent to the students, (2) requires the use of statistical tools to collect and analyze data (activates statistical and technological knowledge), (3) challenges the students to look for solutions to a contextual problem (activates reflective knowledge).

Some tasks presented in real contexts that had potential for introducing, discussing, reflecting and modelling social phenomena were scenarios exclusively used to introduce the calculation of statistical measurements or probabilities. Those were interesting opportunities to promote the development of critical citizenship in students that were lost. For example, a task presented in the context of waste production (see Fig. 17.6) reduced the student’s challenge to transfer information given in a bar graph to a frequency table and to calculate the average production per day.



Fig. 17.5 Type of knowledge promoted by the tasks

The bar graph represents the amount of garbage produced annually by a person in different countries (in kg). Find, approximately, the kilograms of garbage produced by a person in those countries in a day (use calculator): Germany, Japan, Italy, France, and Spain.

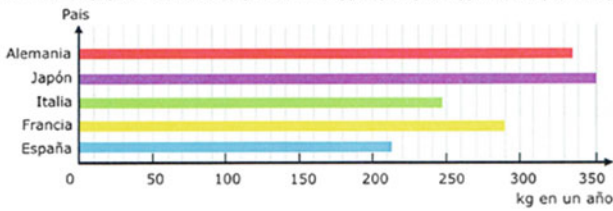


Fig. 17.6 A task in a real context [from *Matemática Experimental 5*, (Uribe-Cálad 2013, p. 249)]

A task with excellent characteristics to model real phenomena and with conditions that could promote reflective knowledge was limited to promoting statistical (averaging) and technological (chart reading) knowledge. A task like this should offer the students an opportunity to discuss the environmental impact of waste production and the ways in which schools, families and individuals could help in reducing the environmental impact. To do that, the task should not focus only on the statistical measures or procedures or in the visual representation of data. Textbook tasks should take students beyond simple calculations and offer opportunities to discuss the context and find and collect relevant data related to the task. Textbooks should ask questions like: What do the numbers say about garbage production in the world?; What can you say about the differences in garbage production between Japan and Spain?; Could you develop a strategy to estimate the amount of garbage your class produces annually?; Can you create a strategy to collect data related to garbage? Could you offer ideas of how to reduce garbage production in your class? This finding is in line with the literature that has found similar results. When tasks use the context of social phenomena, statistical and technological knowledge is privileged over reflection about social or cultural concerns (Stillman et al. 2013).

17.5 Conclusions and Implications

This chapter described the relationship between the statistical components of mathematical textbooks and the development of students' critical citizenship. The main findings reveal that the nature of the tasks proposed in the textbooks is essentially hypothetical; that when the tasks are presented within real contexts, the context is rapidly abandoned to focus exclusively on statistical computations; and that there is a strong emphasis on statistical and technological knowledge over knowledge for social development. The imbalance among the different types of knowledge cannot be attributed exclusively to the textbook. Although the way in which teachers implement the textbook material in the classroom was not the goal of this study, it raises a potential limitation of this research. How the tasks are implemented in the classroom

and how the teacher complements them with current information from students' particular context to promote critical citizenship requires further investigation. Some teachers may add social context when none seems to be present in the task, while others might reduce any social context or reflection—and this would be important to know. It would also be important to know if the findings for grade 5 are true for other grades, particularly in the upper grades.

Perhaps textbooks by themselves fail to prepare students for critical citizenship, but they can become important tools to bring such a curriculum to the classroom. If the tasks presented in the textbooks are inscribed within real contexts—not only to provide scenarios to calculate statistical measures or graphical interpretations—and promote reflective knowledge, teachers might be more likely to help students make connections between statistics and the world and to support the development of socio-political aware citizens.

What was found in relation to the context that accompany the tasks might suggest that the textbooks studied carry conceptions of knowledge anchored in philosophical positions close to Platonism and formalism. In Platonism, mathematics (and other sciences) take place in the world of the ideas where abstraction is privileged. In a similar way, formalism considers mathematics separated from the empirical world (Skovsmose 1999). Consequently, the knowledge that seems to be promoted by the textbooks is often disarticulated, fragmented, inert, rigid, and formal (Makar et al. 2011, Zapata-Cardona and González-Gómez 2017). Perhaps it is time to think about textbooks from philosophical standpoints in which learning is not reduced to abstract and formal knowledge but to develop forms of reflection about the world (Radford 2006).

Acknowledgements This research was supported by The University of Antioquia Research Committee—CODI.

References

- Acevedo-Caicedo, M. M., & Pérez-de-Díaz, M. C. (2013). *Enlace Matemáticas 5* [Mathematics Link 5]. Bogotá: Educar Editores.
- Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, 13(1–2), 5–26. <https://doi.org/10.1080/10986065.2011.538293>.
- Bakker, A., van Mierlo, X., & Akkerman, S. (2012). Learning to integrate statistical and work related reasoning. In *Proceedings from the 12th International Congress on Mathematical Education*. Seoul, Korea.
- Barbosa, J. C. (2006). Mathematical modelling in classroom: A socio-critical and discursive perspective. *ZDM—The International Journal on Mathematics Education*, 38(3), 293–301.
- Beltrán-Beltrán, L. P., Rodríguez-Sáenz, B. P., Suárez-Olarte, A., & Abondano-Mikán, W. (2014). *Contacto Matemático 5* [Mathematics Contact 5]. Bogotá: Editorial Educativa.
- Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3–15). The Netherlands: Kluwer Academic Publishers.
- Biembengut, M. S., & Hein, N. (2002). Modelaje y etnomatemáticas: Puntos(in) comunes [Modeling and ethnomathematics: Points (in) common]. *Números*, 57, 27–39.

- Borba, R., & Selva, A. (2013). Analysis of the role of the calculator in Brazilian textbooks. *ZDM—The International Journal on Mathematics Education*, 45(5), 737–750. <https://doi.org/10.1007/s11858-013-0517-3>.
- Cantor, R. (1997). Los textos de cálculo: Una visión de las reformas y contrarreformas [The texts of calculus: A vision of reforms and counter-reforms]. *Revista EMA*, 2(2), 115–131.
- Chandler, D., & Brosnan, P. (1995). A comparison between mathematics textbook content and a statewide proficiency test. *School Science and Mathematics*, 95(3), 118–123.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly*, 104, 801–823.
- D’Ambrosio, U. (1999). Literacy, matheracy and technocracy: A trivium for today. *Mathematical thinking and learning*, 1(2), 131–153.
- Duran, M. M., Sarmiento, J., Bogotá-Torres, M., Morales, C., & Fuentes, J. (2013). *Matemáticas en Red 5* [Mathematics Network 5]. Bogotá: Ediciones SM.
- Etchegaray, S. C. (2010). Reflexiones y aportes para ayudar a re-pensar la enseñanza de las matemáticas [Reflections and contributions to help re-think the teaching of mathematics]. *Yupana*, 11–23. <https://doi.org/10.14409/yy.v1i5.258>.
- Fan, L., Zhu, Y., & Miao, Z. (2013). Textbook research in mathematics education: Development status and directions. *ZDM—The International Journal on Mathematics Education*, 45(5), 633–646. <https://doi.org/10.1007/11858-013-0539-x>.
- Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, 10(3).
- Giroux, H. A. (1988). *Schooling for democracy: Critical pedagogy in the modern age*. London: Routledge.
- González-Astudillo, M. T., & Sierra-Vasquez, M. (2004). Metodología de análisis de libros de texto de matemáticas. Los puntos críticos en la enseñanza secundaria en España durante el siglo XX [Methodology to analyze mathematics textbooks]. *Revista Enseñanza de las Ciencias*, 22(3), 389–408.
- Guerrero, O. (2008). Educación matemática crítica. Influencias teóricas y aportes [Critical mathematical education: Theoretical influences and contributions]. *Evaluación e Investigación*, 1(3).
- Herbel-Eisenmann, B. A. (2007). From intended curriculum to written curriculum: Examining the “Voice” of a mathematics textbook. *Journal for Research in Mathematics Education*, 38(4), 344–369.
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. Washington, DC: National Center for Education Statistics.
- Howson, G. (2013). The development of mathematics textbooks: Historical reflections from a personal perspective. *ZDM—The International Journal on Mathematics Education*, 45(5), 647–658. <https://doi.org/10.1007/s11858-013-0511-9>.
- Johansson, M. (2003). *Textbooks in mathematics education: A study of textbooks as the potentially implemented curriculum*. (Unpublished doctoral dissertation) University of Lulea, Sweden.
- Jones, K., & Fujita, T. (2013). Interpretations of national curricula: The case of geometry in textbooks from England and Japan. *ZDM—The International Journal on Mathematics Education*, 45(5), 671–683. <https://doi.org/10.1007/s11858-013-0515-5>.
- Joya-Vega, A., Grande-Puentes, X., Acosta, M. L., Ramírez-Rincón, M., Buitrago-García, L., Ortiz-Wilches, L. G., et al. (2014). *Los Caminos del Saber 5* [The Ways of Knowledge 5]. Bogotá: Santillana.
- Lesser, L. M. (2007). Critical values and transforming data: Teaching statistics with social justice. *Journal of Statistics Education*, 15(1). Retrieved from www.amstat.org/publications/jse/v15n1/lesser.html.
- López-Noguero, F. (2002). El análisis de contenido como método de investigación [Content analysis as research method]. *XXI, Revista de Educación*, 4, 167–179.

- Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13(1–2), 152–173. <https://doi.org/10.1080/10986065.2011.538301>.
- Mateos Montero, J. (2008). La “asignaturización” del conocimiento del medio en los textos y contextos escolares. El entorno en las aulas [The “subjectization” of knowledge of environment in the texts and school contexts. The environment in the classrooms]. *Investigación en la Escuela*, 65, 59–70.
- Mejía-Fonseca, C. F., Guzmán-Pineda, L. E., Vega-Reyes, A. M., & Baquero-Guevara, D. C. (2011). *Zona Activa 5* [Active Zone 5]. Bogotá: Voluntad.
- Mesa, V. (2000). *Conceptions of function promoted by seventh- and eighth-grade textbooks from eighteen countries*. University of Georgia: Unpublished doctoral dissertation.
- Nie, B., Freedman, T., Hwang, S., Wang, N., Moyer, J., & Cai, J. (2013). An investigation of teachers’ intentions and reflection about using standards-based and traditional textbooks in the classroom. *ZDM—The International Journal on Mathematics Education*, 45(5), 699–711. <https://doi.org/10.1007/s11858-013-0493-7>.
- Österholm, M., & Bergqvist, E. (2013). What is so special about mathematical text? Analyses of common claims in research literature and of properties of textbooks. *ZDM—The International Journal on Mathematics Education*, 45(5), 751–763. <https://doi.org/10.1007/s11858-013-0522-6>.
- Otte, M. (1986). What is a text? In B. Chistiansen, A.G., Howson, & M. Otte (Eds.). *Perspectives on mathematics education* (pp. 173–203). Dordrecht: D. Reidel Publishing Company.
- Pepin, B., Gueudet, G., & Trouche, L. (2013). Investigating textbooks as crucial interfaces between culture, policy and teacher curricular practice: Two contrasted case studies in France and Norway. *ZDM—The International Journal on Mathematics Education*, 45(5), 685–698. <https://doi.org/10.1007/s11858-013-0526-2>.
- 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). The Netherlands: Kluwer Academic Publishers.
- Pfannkuch, M., Budgett, S., Fewster, R., Fitch, M., Pattenwise, S., & Wild, C. (2016). Probability modeling and thinking: What can we learn from practice? *Statistics Education Research Journal*, 15(2), 11–37.
- Prendes-Espinosa, M. (2001). Evaluación de materiales escolares [Assessment of educational materials]. *Revista Píxel Bit*, 16, 1–20. <http://www.sav.us.es/pixelbit/pixelbit/articulos/n16/n16art/art167.htm>.
- Radford, L. (2006). *Elementos de una teoría cultural de la objetivación* [Elements of a cultural theory of objectification] (pp. 103–129). Número Especial: Revista Relime.
- Radford, L. (2016). On alienation in the mathematics classroom [Sobre la alineación en la clase de matemáticas]. *International Journal of Educational Research*, 79, 258–266. <https://doi.org/10.1016/j.ijer.2016.04.001>.
- Remillard, J. T. (2005). Examining key concepts in research on teachers’ use of mathematics curricula. *Review of Educational Research*, 75(2), 211–246.
- Roth, W.-M. (1996). Where is the context in contextual word problems? Mathematical practices and products in Grade 8 students’ answers to story problems. *Cognition and Instruction*, 14, 487–527.
- Salcedo, A. (2015). Exigencia cognitiva de las actividades de estadística en textos escolares de Educación Primaria [Cognitive demand of statistical activities in primary school textbooks]. In J. M. Contreras, C. Batanero, J. D. Godino, G. R. Cañadas, P. Arteaga, E. Molina, et al. (Eds.), *Didáctica de la Estadística, Probabilidad y Combinatoria. Actas de las Segundas Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 307–315). Granada.
- Shimizu, Y., Kaur, B., Huang, R., & Clarke, D. J. (2010). *Mathematical tasks in classrooms around the world*. Rotterdam, The Netherlands: Sense Publishers.
- Silva-Calderon, L. H. (2014). *Avanza Matemáticas 5* [Advancing Mathematics 5]. Bogotá: Norma.

- Skovsmose, O. (1992). Democratic competence and reflective knowing in mathematics. *For the Learning of Mathematics*, 12(2), 2–11.
- Skovsmose, O. (1999). *Towards a philosophy of critical mathematics education*. (P. Valero, Trad.) Bogotá: Una empresa docente.
- Stillman, G., Brown, J., Faragher, R., Geiger, V., & Galbraith, P. (2013). The role of textbooks in developing a socio-critical perspective on mathematical modeling in secondary classrooms. In G. A. Stillman (Ed.), *Teaching mathematical modelling: Connection to research and practice. International perspectives on the teaching and learning of mathematical modelling* (pp. 361–371). Dordrecht: Springer Science + Bussiness. https://doi.org/10.1007/978-94-007-6540-5_30.
- Travé González, G., & Pozuelos Estrada, F. (2008). Consideraciones didácticas acerca de las líneas de investigación en materiales curriculares. A modo de presentación [Didactic considerations about research in curricular materials. A presentation]. *Investigación en la Escuela*, 65, 3–10.
- Uribe-Cálad, J. A. (2013). *Matemática Experimental 5* [Experimental Mathematics 5]. Medellín: Uros Editores.
- Usiskin, Z. (2013). Studying textbooks in an information age—A United States perspective. *ZDM—The International Journal on Mathematics Education*, 45(5), 713–723. <https://doi.org/10.1007/s11858-013-0514-6>.
- Valero, P. (2002). Consideraciones sobre el contexto y la educación matemática para la democracia [Considerations about the context and the mathematics education for democracy]. *Cuadrante*, 11(1), 49–59.
- Xu, B. (2013). The development of school mathematics textbooks in China since 1950. *ZDM—The International Journal on Mathematics Education*, 45(5), 725–736. <https://doi.org/10.1007/s11858-013-0538-y>.
- Zapata-Cardona, L., & González-Gómez, D. (2017). Imágenes de los profesores sobre la estadística y su enseñanza [Teachers' Images about Statistics and its Teaching]. *Educación Matemática*, 29(1), 61–89. <https://doi.org/10.24844/EM2901.03>.

Chapter 18

Critical Mathematics Education and Statistics Education: Possibilities for Transforming the School Mathematics Curriculum



Travis Weiland

Abstract This chapter discusses how ideas from critical mathematics education and statistics education intersect and could be used to transform the types of experiences that students have with both mathematics and statistics in the school mathematics curriculum. Key ideas from the critical mathematics literature are described to provide a background from which to discuss what a critical statistics education could be. The chapter ends with a discussion of some of the major barriers that need to be considered to make such a vision a reality and possible future directions for moving towards making a critical statistics education a reality.

Keywords Critical literacy · Critical mathematical education · Statistics education
Statistical literacy

18.1 Importance of Statistics in School

Data are everywhere in society today, aimed at influencing our decisions about what toothpaste to buy, what politician we should vote for, or what medicine is the best treatment for what ails us (Steen 2001). Today huge breakthroughs in science, medicine, economics, and public policy are being made using advanced data modeling techniques (Davidian and Louis 2012). Statistics, which is often described as the science of data, is becoming an increasingly important topic of study because of our society's reliance on data (Ben-Zvi and Garfield 2008; Gattuso and Ottaviani 2011). Experts and policy makers in many fields are increasingly basing their decisions on statistical results, using data to draw out new insights about the world (Pfannkuch 2008). In data driven societies, it is crucial that individuals are able to interpret and critically analyze quantitative data and statistics (Ben-Zvi and Garfield 2004) to be critical citizens (Skovsmose and Valero 2008). As the need has increased for governments to become more transparent in their operation and decision making,

T. Weiland (✉)

Department of Mathematical Sciences, Appalachian State University, Boone, NC, USA

e-mail: weilandtj@appstate.edu

© Springer Nature Switzerland AG 2019

G. Burrill and D. Ben-Zvi (eds.), *Topics and Trends in Current Statistics Education Research*, ICME-13 Monographs, https://doi.org/10.1007/978-3-030-03472-6_18

it has become crucial for citizens to have a strong statistical literacy to make sense of technical reports (Ullmann 2016). Furthermore, recent votes in both England (Brexit) and the United States (2016 Presidential Election) were fraught with misinformation campaigns many of which included the misuse of data based arguments (Belham 2016), which points to the importance of statistical literacy for being an engaged critical citizen. One of the commonly held goals of public education in the United States is to prepare students to become citizens of society (Labaree 1997). In light of the data centric focus of modern societies, a goal of public K-12 education should include teaching students to be statistically literate active citizens in their data-driven societies.

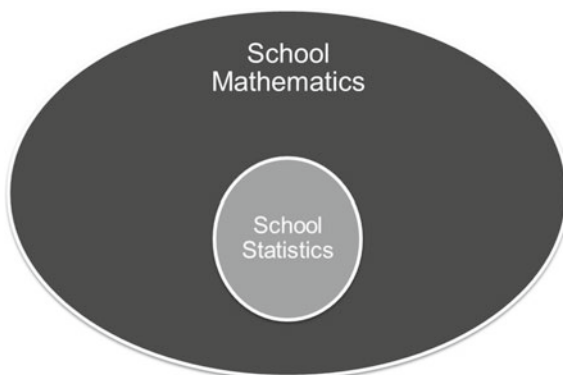
As the prevalence of discourse using data increases it is important that students are able to critically interpret such discourses in their context in society. The term critical is used here not in the critical thinking sense that is generally used in education drawing upon the traditional ways of knowing of different disciplines. Instead it is used in the way critical or emancipatory pedagogical writings describe to interrogate, problematize, and reconstitute discourses that are dehumanizing, unjust, and position groups as others (Darder 2014; Freire 1970; Giroux 2011; Gutstein 2003). In the case of statistics this involves using statistics to investigate the underlying structures and hidden assumptions present in society and also to critique and understand the hidden assumptions in the use of statistics. Many assumptions come with the use of quantitative data that are not always apparent in the contexts students are given to investigate in their school classrooms.

This chapter presents an argument, rooted in past scholarship, for a critical statistics education in conjunction with critical mathematics education (Skovsmose 1994a) in K-12 education with the intention of giving students experiences with reading and writing the world with mathematics (Gutstein 2003, 2006, 2013) and statistics. This is not an entirely new idea as Lesser (2007) has introduced the notion of teaching statistics with social justice, and a number of scholars have begun to delve into the complexities of raising sociopolitical issues in statistics classes (e.g. Bergen 2016; Engel 2016; Gray 2016; Poling and Naresh 2014). Also Gal (2002) began to draw Freire's (1970) ideas into his work on statistical literacy more than a decade ago in terms of critical dispositions. However, the majority of this work has been done or written for a post-secondary statistics education audience, and only a small subset of most societies' citizens goes on to post-secondary education. The use of data based arguments and statistics today is exploding in almost every setting making it crucial for all individuals to have experiences to learn and use statistical concepts and practices to make sense of the types of sociopolitical issues they will need to navigate and make sense of in their daily lives as citizen's in today's modern societies.

18.2 The Complex Nature of Statistics in the School Setting

K-12 educational settings are very different and far more complex than most post-secondary settings in several regards. To begin with students in K-12 settings are not yet considered adults and are typically under the direct care of their parents

Fig. 18.1 The situatedness of statistics in school



and families in many nations, which also means that families also commonly have a voice in their children's education. Past research has shown parents can be quite influential in shaping what students are taught in mathematics classrooms (Boaler 2002; Herbel-Eisenmann et al. 2006). Unlike most post-secondary settings where students predominantly self-select the institution and curriculum in which they wish to participate, students in K-12 settings have little agency over the institutions they attend or the curriculum they experience, which is predominantly determined by geography and governmental policy. Furthermore, what students are taught in schools must be negotiated by a significant number of stakeholders including: parents, policy makers, teachers, researchers, disciplinary experts, politicians, etc. (Apple 1992). These stakeholders in turn bring all their own beliefs, values, and perspectives to bear on shaping the school mathematics curriculum. Unfortunately, the result has often been a very neutral mathematics curriculum, which does not address the truly political nature of mathematics education or the formatting power it has in shaping the world around us (Gutiérrez 2013a; Skovsmose 1994b).

The teaching and learning of statistics in the school setting is further complicated because it is situated in the mathematics curriculum, as modeled in Fig. 18.1, where it has only begun to gain a foothold in terms of statistical thinking and reasoning (Scheaffer and Jacobbe 2014). Furthermore, statistics is generally taught by mathematics teachers who may have had little to no past experience with statistics (Shaughnessy 2007), not by statisticians who generally teach much of post-secondary statistics. This means that many of the teachers directly shaping the curriculum that students experience in the mathematics classroom are likely more enculturated in the practices of the discipline of mathematics than they are of the discipline of statistics. Unfortunately, as Eichler and Zapata-Cardona (2016) point out, empirical research on mathematics teachers' teaching of statistics in the K-12 setting has been limited up to this point. Pointing out some of the complexity of considering statistics education in the K-12 school setting illustrates how different it is in many regards compared to post-secondary settings where much of the statistics education work around sociopolitical issues has been done.

18.3 Critical Mathematics

The world today is faced with a multitude of challenges such as economic collapse, poverty, resource depletion, climate change, polarization in wealth, extreme nationalism, and immigration/migration. As a result more and more scholars are advocating for bringing these issues into school classrooms (Apple and Beane 2007; Giroux 2011; Ladson-Billings 1995). Scholars in mathematics education have been advocating for similar efforts in the teaching of mathematics as well. In consideration of the social, political, and ethical dimensions of mathematics education, scholars in mathematics education over the past two decades have begun calling for the use of critical (Frankenstein 1994; Skovsmose 1994a; Wager and Stinson 2012) and culturally relevant pedagogies (Gutstein et al. 1997; Ladson-Billings 1995). These scholars seek to create mathematics classrooms where students learn how to understand their social, cultural, and political context in society as well as how to change that context. There has also been a growing literature base in the field of mathematics education based around incorporating social and political critique into the mathematics curriculum (Gutiérrez 2009, 2013a; Skovsmose 1994a; Skovsmose and Valero 2008). Many of these scholars argue for centering pedagogy around problem posing and connecting content areas to fundamental questions of society rather than focusing on neutral or trivial problems or contexts (Frankenstein 2009; Freire 1970; Gutstein 2006; Gutstein and Peterson 2013). This is a serious problem because as Skovsmose (1994b) points out,

It is important to relate the idea of the invisibility of mathematics to the assumption about the formatting power of mathematics, because if both assumptions are correct, we witness a challenging and critical situation for mathematics education. This conflict has been formulated as the paradox of relevance: on the one hand, mathematics has a pervasive social influence and, on the other hand, students and children are unable to recognize this relevance. (p. 82)

Critical mathematics scholars argue that students need opportunities to see and experience the pervasive influences mathematics has on the social world. To clarify how the word critical is used in this scholarship, Gutstein et al. (1997) bring up an interesting point about critical mathematical thinking and how critical has two meanings in this instance. One meaning for critical is in the mathematical sense as in making sense of problems, creating arguments, making conjectures, critiquing the reasoning of others, ideas that generally fall under the term critical thinking. We see these in NCTM's (2000) *Principles and Standards*, which have been taken up in a wide variety of educational settings worldwide. For example, the reasoning and proof standard includes “make and investigate mathematical conjectures” and “develop and evaluate mathematical arguments and proofs” (NCTM 2000, p. 56). There is also the meaning of critical in the broad sense, using multiple perspectives to look at an issue, and questioning the context in which one is situated and in education questioning standards, curriculum and practices (Gutstein et al. 1997). It is this second meaning of critical that critical mathematics education scholarship contributes to mathematics

education. As Skovsmose (1994b) describes in his book *Towards a Philosophy of Critical Mathematics Education*,

If educational practice and research are to be critical, they must address conflicts and crises in society. Critical education must disclose inequalities and suppression of whatever kind. A critical education must not simply contribute to the prolonging of existing social relationships. It cannot be the means for continuing existing inequalities in society. To be critical, education must react to the critical nature of society. (p. 22)

In Gutstein's (2003, 2006) writings he discusses his work teaching mathematics for social justice in urban public schools incorporating issues brought by his students such as urban planning, stop and frisk, gentrification, and AIDs/HIV. He discusses how he created challenging mathematics curriculum with students that pushed students to be academically successful in mathematics but also provided students with experiences using mathematical concepts to investigate and critique their own context in society. Gutstein (2006) draws heavily from Paulo Freire's literacy work, which was done predominantly in Brazil to help the marginalized of that nation in the latter half of the twentieth century to become literate, to make sense of and in turn influence and improve their reality and position in their world. Freire and Macedo (1987) discussed literacy in terms of reading the word and the world, learning to make sense of symbol systems by using them in conjunction with making sense of the world around oneself. They also discussed writing the word and the world, emphasizing how literacy can empower people to make sense of the world around them but to also influence and shape the world around them. Gutstein (2006) draws heavily from these notions to describe how to envision reading and writing the world with mathematics. He describes reading the world with mathematics as meaning:

to use mathematics to understand relations of power, resource inequities, and disparate opportunities between different social groups and to understand explicit discrimination based on race, class, gender, language, and other differences. Further, it means to dissect and deconstruct media and other forms of representation. It means to use mathematics to examine these various phenomena both in one's immediate life and in the broader social world and to identify relationships and make connections between them. (Gutstein 2003, p. 45)

This definition emphasizes how mathematical literacy can be used to read the word, which increasingly includes mathematical and quantitative language (Steen 2001) and also to read the world, which has been structured based on quantitative and technological discourses rooted in the abstract language of mathematics (Skovsmose 1994b). Reading the world with mathematics can in turn lead to writing the world with mathematics, which Gutstein defines as:

using mathematics to change the world... I view writing the world with mathematics as a developmental process, of beginning to see oneself capable of making change, and I refer to writing the world for youth as developing a *sense of social agency*. A "sense" of social agency captures the gradual nature of students' growth-it is not an all-or-nothing proposition. (Gutstein 2006, p. 27)

Writing the world with mathematics in this sense also implies being able to use mathematics in a meaningful way for positively changing the world, which would

seem to be very much in line with the goals of mathematics curriculum policy documents such as the Common Core Standards for Mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers 2010) in the United States. However, in that document's description of its goal for mathematics education, making every student college and career ready, there is no specific connection made to changing the world for the better. It is neutral with respect to ethics to guide what constitutes a positive change or to even consider positive change in the context of mathematics, which is something that critical literacy emphasizes (Giroux 1993) and that can be seen in Gutstein's (2003, 2006) definitions of reading and writing the world with mathematics.

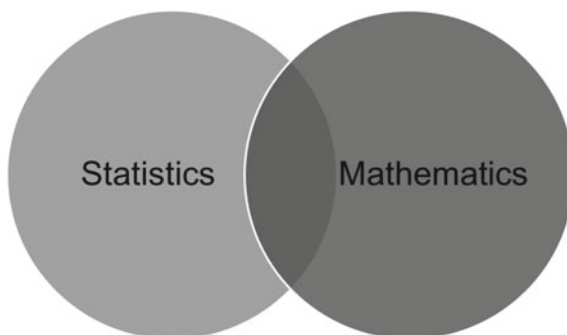
Mathematics can be used to change the world in very unjust ways. For example, consider the recent economic recession after the housing bubble burst. A number of bankers were blamed for using their abilities with mathematics, in unscrupulous ways, to cheat individuals out of large sums of their money, which sent the economy into a downward spiral (Cohan 2015). Just because an individual has the mathematics skills and abilities to be ready for a career does not mean they have any ethical principles behind how to use those abilities. If the goal of education truly is to prepare critical citizens for participation in democracy (Apple and Beane 2007; Giroux 1989) there needs to be some sense of developing ethics to guide such participation (Giroux 1993). Ethics is often linked to notions of fairness and social justice in mathematics education (Boylan 2016) but is a very murky term used and taken up in a wide variety of ways. I use the term ethics here broadly to mean a set of philosophical or moral values that people use to make decisions in relation to others, their communities, and the world at large.

The notions of critical mathematics and the idea of reading and writing the word and the world with mathematics can be drawn on in considering how to in turn foster students in K-12 school settings to read and write both the word and the world with statistics. Critical statistics education is needed in spite of a growing literature base around critical mathematics education because, as described in the next section, the disciplines of mathematics and statistics are distinct.

18.4 The Disciplines of Mathematics and Statistics

Mathematics is a socially created dynamic body of knowledge with a social history, and with areas expanding and contracting over time situated in context (Bishop 1988; Davis and Hersh 1981). A common definition of mathematics drawing from the Oxford English Dictionary is "the abstract science of number, quantity, and space" (Mathematics n.d.). However, what is considered to be the specific scope and knowledge of mathematics is socially agreed upon by members of the discipline and changes with time (Davis and Hersh 1981), which makes it inherently political. As a point of clarification, I am using the term political in the way it is commonly

Fig. 18.2 The relationship between the disciplines of mathematics and statistics



used in sociopolitical perspectives, which is to describe any situation that involves making a choice or decision as there are always multiple options or perspectives in such situations, where one is chosen and advantaged over others and such decisions are always situated in power relations. From the sociopolitical perspective defining mathematics is political in the sense that its boundaries, practices, and what is considered knowledge are situated in, and shaped by, power relations between individuals and institutions, where certain views and perspectives may advantage some, while disadvantaging or silencing others (Gutiérrez 2013b). Statistics formally came into fruition in the 18th century (Katz 2009; Stigler 1986) and has similar characteristics. However, instead of being looked at as the science of quantity and space it is often viewed as the science of data or measurement (Davidian and Louis 2012; Stigler 1986). Both mathematics and statistics are part of the mathematical sciences (Steen 2001). As Moore and Cobb (2000) describe, “mathematicians and statisticians share a commitment to a process of pattern searching, generalization, and verification that operates at a deep level, despite surface differences” (p. 22). Furthermore, mathematics and statistics are linked through probability, which is a part of mathematics that is crucial to statistical inference (Fienberg 1992). However, statistics is its own distinct discipline not a sub-discipline or branch of mathematics (Cobb and Moore 1997; Franklin et al. 2007; Gattuso and Ottaviani 2011; Groth 2013), but it is inextricably linked to mathematics as modeled in Fig. 18.2. As Steen (2001) points out, “Although each of these subjects shares with mathematics many foundational tools, each has its own distinctive character, methodologies, standards, and accomplishments” (p. 4).

Some of the main differences between statistics and mathematics are discussed in the literature and are described in the next section. This is done not to create divisions or to argue for taking statistics out of the mathematics curriculum but instead to embrace the diversity of these two disciplines and how they provide different ways of looking at the world. Furthermore, such differences can position statistics as a powerful entry point for interrogation of sociopolitical issues in the school mathematics curriculum, which is discussed later.

18.5 Differences Between Statistics and Mathematics

There is a strong literature base that discusses the differences between the discipline of mathematics and that of statistics (Cobb and Moore 1997; Franklin et al. 2007; Gattuso 2006; Gattuso and Ottaviani 2011; Groth 2007, 2013; Pfannkuch 2008). The major differences discussed in these works are context, variability, uncertainty, and inductive versus deductive reasoning.

In statistics, “data are not just numbers they are numbers with a context” (Cobb and Moore 1997, p. 801). This is a departure from mathematics, particularly what is taught in schools, where numbers are frequently presented and used in their abstract form without any connection to context (Gattuso and Ottaviani 2011). As Cobb and Moore (1997) discuss, in mathematics, context is generally stripped away from a problem to try and uncover, or abstract, the underlying mathematical structure of the context. However, in statistics, the analysis of data cannot be considered without thinking about the context of the data (Cobb and Moore 1997; Franklin et al. 2007; Wild and Pfannkuch 1999). There is a constant interplay between considering a statistical problem and the context of the problem (Groth 2007; Wild and Pfannkuch 1999).

The need for the discipline of statistics comes from the omnipresence of variability (Cobb and Moore 1997) in the world. Simply stated in most cases individuals or objects that we study are not all the same for every attribute. Therefore statistics focuses on how attributes can vary from individual to individual or object to object. In statistics there are four main kinds of variation: measured, natural, induced, and sampling (Franklin et al. 2007).

It is important to point out that variation is not absent from mathematics, but that it is considered in a very deterministic way. Consider a linear function, the values of dependent and independent variables covary with one another but in a specific unchanging way (e.g. “as x increases by 2 y increases by 3”). The discipline of statistics uses the concept of linear functions. However, instead of determining what the value of the dependent (y) variable will be, given the value of the independent (x) variable, the function is used to “fit” the data. In other words, a linear function can be used to summarize the relationship between an independent or explanatory variable and a dependent or response variable, where the accuracy of a prediction depends on the amount of variation between the observed and predicted values of the response variable explained by the explanatory variable in the model. This procedure is used in an effort to create a linear function that best “fits” the data, but will only provide predictions, which may not be very accurate depending on the amount of variation the model can account for in the actual data values.

As a result of the omnipresence of variation in statistical investigations, there is no certainty in the solutions. The end product of a statistical investigation is better thought of as a well principled argument (Abelson 1995). Mathematics on the other hand is generally treated in a very deterministic way, logically deducing a single solution to a problem using theorems, axioms, and definitions from the community of mathematics (Gattuso and Ottaviani 2011).

Another main difference between statistics and mathematics is the type of reasoning generally used. Mathematics primarily relies on deductive reasoning using definitions, axioms, and theorems, in a logical chain of reasoning, to come to a conclusion. For example, a student could use Euclid's definition of a circle and his first and third postulates to construct an equilateral triangle. At the same time Euclidean geometry is based on certain unprovable assumptions such as the parallel postulate, which if changed creates an entirely new type of geometry and way of viewing the world (Katz 2009). The practice of statistics is often driven by a question for which data are collected analyzed and interpreted to answer the question (Franklin et al. 2007; Wild and Pfannkuch 1999). It is from the data that information is empirically derived, which is the hallmark of inductive reasoning. Similar to uncertainty this can lead to issues in teaching statistics as teachers who have had few experiences with statistics may attempt to deduce solutions from rules and assumptions to find a single certain answer rather than inducing them from the data to find a range of possibilities.

These differences can also lead to common statistics teaching practices that are different from those of mathematics. In this regard, teaching concepts from statistics is not the same as teaching concepts from mathematics, though clearly there are parallels between the two (Franklin et al. 2007; Gattuso 2006; Gattuso and Ottaviani 2011). Since statistics is often situated within the mathematics curriculum at the K-12 level, this position can give the impression that statistics is just a branch of mathematics (Groth 2007). This is not to say that statistics should be taught outside of mathematics at the K-12 level (Franklin et al. 2007; Usiskin 2014). However, it is important to understand that statistics is a distinct discipline and as such there are different strategies, habits of mind, and practices involved in teaching concepts from statistics (Cobb and Moore 1997; Groth 2007). One approach is that of Gattuso and Ottaviani (2011) who aim "to emphasize the necessity of complementing statistical thinking and mathematical thinking in school and generating didactic strategies allowing statistics and mathematics to evolve together, in a harmonious way" (p. 122). They also state that statistics concepts and problems can be used to compliment mathematical thinking and bring more context and students' interests into mathematics classrooms. For these topics to evolve together in K-12 curriculum teachers should know both their similarities and their differences. How critical statistics could be envisioned differently than critical mathematics based on some of the disciplinary differences described in this section is elaborated on in the next section.

18.6 Differences Between Critical Statistics and Critical Mathematics

Critical mathematics has had several decades to build a base of literature and create examples for using mathematics for social justice. A number of the examples of critical mathematics activities in the classroom involve concepts from statistics. This

section presents two examples of such activities and points out how these activities could be expanded to provide students with robust experiences involving statistics.

Skovsmose and Valero (2008) in their paper on *Democratic access to mathematically powerful ideas* use an example that involves drawing samples of eggs from a population of eggs and seeing how many are infected based on a rate of infection reported by the Dutch government. This is inherently a statistical task creating a model based on chance to model getting an egg infected with salmonella. Some of the questions the authors had the students consider as part of this task were:

The basic question to be addressed by this experiment has to do with the *reliability* of information provided by samples. How can it be that a sample does not always tell the “truth” about the whole population? And how should we operate in a situation where we do not know anything about the whole population, except from what a sample might tell? How can we, in this case, evaluate the reliability of numerical information? (p. 9)

These questions are very rooted in statistical practices. This activity discussed by Skovsmose and Valero is meant to begin to get students to discuss the differences between ideal mathematical calculations and figures from empirical data collection, which is an important difference between the ways of knowing in mathematics and those of statistics. However though sampling was discussed, there was no emphasis on a discussion of variability, which would become a focal point if there had been an explicit focus on teaching statistics in this activity. For example, students could compare their samples across groups to facilitate a discussion of sampling variability, which could be expanded upon to begin to introduce the idea of sampling distributions as well as develop the ideas of standard error and margins of error to emphasize the variability present in empirical work.

Again the point is not to separate mathematics and statistics as they are deeply connected and statistics relies heavily on mathematics (Groth 2007, 2013). Instead, the aim is to point out the differences so that they are emphasized and not lost in classroom mathematics instruction. For example, in Skovsmose and Valero’s (2008) discussion of the egg task, they move into the mathematical realm of calculating theoretical probabilities instead of highlighting the idea of variation and how it can be measured and interpreted. Now this move might be because Skovsmose and Valero are trying to communicate to the mathematics education community in this particular example. However, in many cases this is the same community that is tasked with teaching statistics in the school setting and could potentially benefit from seeing specific examples of statistical practices in conjunction with the teaching of mathematics.

Another example comes from a project that Gutstein (2003) used in one of his mathematics classes described here:

For example, I developed a project in which students analyzed racially disaggregated data on traffic stops. The mathematical concepts of proportionality and expected value are central to understanding racial profiling. Without grasping those concepts, it is hard to realize that more African American and Latino drivers are stopped than one would expect, and this disproportionality should lead one to examine the root causes of the anomaly (p. 49).

Similar to the egg problem, the mathematical concepts become the focus of the discussion where there are also opportunities to discuss and investigate important

statistical concepts. In this example Gutstein (2003) focuses on the idea of proportionality to discuss how the African American and Latino/a drivers were pulled over disproportionately due to probabilistically determined expected values. However, these disaggregated data can be considered an example of a sample, which provides an opportunity to talk about variation in samples and to look at this proportion in terms of a sampling proportion. A sample proportion has a margin of error that needs to be considered when making inferences to a population. This context also provides the opportunity to introduce the idea of sampling distributions, which could be developed using a simulation to bootstrap a sampling distribution from the sample drawn. In this way, students could also begin to investigate what it means for a sample proportion to be unlikely to be drawn due to chance and to begin to make inferences about the population proportion.

The discussion above illustrates how rich mathematical lessons could be used to teach statistical concepts in mathematics classrooms to possibly help students to read and write both the word and the world with both mathematics and statistics. It is important that statistics educators begin to contribute more to these conversations based on their own expertise, which could be used to begin to emphasize statistical and mathematical concepts and practices in school curriculum.

18.7 Critical Statistics Education

In this section I outline a possible vision for critical statistics education beginning by briefly drawing a broad connection between Freire's (Darder 2014; Freire 1998; Freire and Macedo 1987) notions of literacy and Gal's (2002) description of statistical literacy in conjunction with statistical enquiry (Franklin et al. 2007; Wild and Pfannkuch 1999) to situate a view of critical statistics education in an overarching literacy perspective. The remainder of the discussion focuses on elaborating several points of intersection that connect the ideas discussed in earlier sections relative to critical mathematics education to statistics education, namely: considering context, variation, subjectivity, transnumeration, and problem posing.

18.7.1 Literacy Perspective

The term statistical literacy has been used by many scholars with many different meanings attributed to it (Ben-Zvi and Garfield 2004). In this chapter I have chosen to draw from Gal's definition of statistical literacy because of its seminal importance and because it is still one of the most commonly used definitions. Gal's (2002) states:

The term *statistical literacy* refers broadly to two interrelated components, primarily (a) people's 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) their ability to *discuss or communicate* their reactions to such statistical infor-

mation, 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. 49).

In his discussion, Gal (2002) describes two different types of contexts. The first, which is the focus of his definition of statistical literacy, he refers to as the reading context, which he describes as “people’s ability to act as effective “data consumers” in diverse life contexts” (p. 50). The other is the enquiry context, which is described as, “in enquiry contexts individuals serve as ‘data producers’ or ‘data analyzers’ and usually have to interpret their own data and results and report their findings and conclusions” (p. 50). A connection can be made here to the idea of reading and writing the world with mathematics (Gutstein 2006), discussed earlier. I propose that reading the world with statistics is what Gal (2002) describes as the reading context in statistical literacy and writing the world with statistics is partially what Gal (2002) refers to as enquiry contexts. I say partially because writing the world refers to changing one’s context, which goes beyond just producing and analyzing data to using it to take action to change the context it describes. Gal’s definition of statistical literacy focuses on the reading context, though he does mention the enquiry context. To further elaborate on the enquiry context to describe writing the world with statistics from the critical literacy perspective, I am choosing to draw on the statistical investigative cycle from the GAISE framework (Franklin et al. 2007) and from Wild and Pfannkuch’s (1999) work on statistical enquiry.

18.7.2 Context

One aspect of statistics that makes it particularly powerful for dealing with issues of race, gender, sexuality, immigration, sustainability and other sociopolitical issues is that it is the science of data, and data are inherently situated in context. Statistics can in fact be a gateway to introducing contextual discussions, situated in the daily realities of students, into the mathematics classroom, and in this way it can serve to act as a lens for reading the world. Statistics helps provide tools, practices, and habits of mind to measure and make sense of patterns in the world around us. It is this very focus that makes it so powerful for reading the world.

In relation to statistics education this perspective requires that teaching and learning statistics is situated in actual meaningful contexts for students. Frankenstein (2009) in her work discusses the importance of real real-world problems in mathematics education, where the context does not just serve as “window dressing” for a mathematics problem, but the actual focus and purpose of the problem is to explore and learn more about a context. From a disciplinary standpoint statistics is aptly suited to take on this task. However, to do so the issue must be taken head on in the school setting where mathematics and statistics are often portrayed as neutral, situated in fictitious contextualized situations (Frankenstein 2009; Skovsmose 1994a). This means contexts such as people’s favorite color, age, or height should not be used

as window dressing for tasks focused on calculation or following routine algorithms; instead instruction should focus on contexts that are from students daily lives and the contexts in which they are situated and using those as spring boards for students to explore and learn statistics in practice as part of making sense of a meaningful context.

For example, when looking at a data set with categories of male and female for gender a discussion around what is gender could begin. A discussion could revolve around whether these two categories are adequate for capturing the gender diversity of a population or as some argue whether they really exist at all except as a social construct (Butler 1990). This discussion could also move from reading context to enquiry context (Gal 2002) discussing issues of how to collect data on a person's gender to actively investigate issues around gender identity, which could then be interpreted and reported to write the word. Could data be collected by allowing an individual to self-identify using a fill in the blank item versus a dichotomous choose-one item? What implications does such a choice have on how the data can be analyzed and interpreted? Such investigations might also be focused on identifying and uncovering issues of genderism related to access to education, wages, or representation in government and society in an effort to transform conditions to write the world. Discussions of gender and other social constructs also relate to the notion of operational definitions of social constructs, which is an issue specific to statistics, which should be part of critical statistics education and has been discussed by others (e.g. Lesser 2007).

Drawing in such context also inherently brings in other issues particularly at the K-12 level where mathematics has traditionally been taught in a very neutral form (Gutiérrez 2013a; Skovsmose 1994b). Bringing sociopolitical contexts common in today's modern societies into the classroom also means opening up the classroom to the divisive and at times very insensitive and confrontational discourse that is also prevalent in societies around such issues. This raises a number of issues for the implementation of such curriculum that will be discussed and elaborated on later in the chapter.

18.7.3 *Variability*

One of the main purposes for the discipline of statistics is in attending to the omnipresence of variability inherent in our world (Cobb and Moore 1997). It is this very focus that makes statistics so powerful for reading the world. Variability comes in many forms, from how we measure things, to how we sample things, to how we try to show "cause and effect" by creating conditions to induce variability, to the fact that populations of things vary in measure of their attributes from one thing to the next. The word things is appropriate because statistics really is that broad in its application that it is applied to living things (people, animals, plants, etc.) and inanimate objects both created by people (machines, products, emissions, etc.) and created naturally (rocks, planets, stars, geological formations, etc.). In statistics there is an explicit emphasis

on trying to make sense of variability, measure it, visualize it, and at times control it. In the educational setting, explicit acknowledgement and treatment of variability is crucial (Cobb and Moore 1997; Franklin et al. 2007; Shaughnessy 2007), not only because of its central role in the discipline, but also because it helps to provide students with the perspective that everything does not always fit into a single numeric answer.

18.7.4 Subjectivity

Another issue to consider is how people's subjectivity plays a role both in the production of data based arguments and in the interpretation of such arguments. This relates to reading and writing both the word and the world as our subjectivities influence everything we see, say, do, and make sense of. Our subjectivities act to filter how we can experience and act upon the world (Foucault 1972; Gutiérrez 2013b; Harding 1991). In statistics education, subjectivity is often treated in relation to biases in data production, such as how survey questions are worded, samples chosen, and participants or things are assigned to groups (Franklin et al. 2007; Utts 2003). However, what is not always considered is how the subjectivity of the reader of data based arguments plays a role in the reader's interpretation of the arguments. For example, consider the current anti-vaccination movements that are growing in spite of overwhelming amounts of scientific data that support the benefits of vaccination and the lack of scientific data to support claims of the purported negative effects of vaccinations. Some of my own subjectivities are likely evident in this statement and also throughout this chapter. People's subjectivities are always present in making sense of statistical arguments or carrying out statistical investigations whether they are lurking below the surface or transparently made explicit. Therefore it is crucial that authors make their subjectivities explicit in their data based arguments and that they reflect on how such subjectivity might influence their arguments.

There is a reason why "alternative facts" can spread like wild-fire in spite of a lack of supporting evidence, and it relates to the subjectivities through which people filter the world, making some statements more plausible, while others less. Open discussions of such issues in the classroom can prepare students to make sense of such alternative facts and should be a part of a critical statistics education.

18.7.5 Transnumeration

The construct of transnumeration comes from past research in statistics education on the types of statistical thinking involved in statistical enquiry and is defined as "numeracy transformations made to facilitate understanding" (Wild and Pfannkuch 1999, p. 227). A number of different transformations of representations are involved in modeling data including the initial measurement of some real-world phenomenon,

applying aggregate measures to represent data, constructing data visualizations to represent data, creating statistical arguments that are convincing and understandable to communicate to an intended audience related to the problem situation being investigated (Pfannkuch et al. 2002). Transnumeration is related to context, variability, and subjectivity that were discussed previously. It is related to context as the initial operationalization and measurement of things involves transnumeration by quantifying and classifying the things, changing their representation to better understand them. This is also related to subjectivity because it is the person organizing and carrying out a statistical investigations that determines how to represent reality through how they decide to classify and quantify the aspects of reality they are measuring.

The focus on variability also connects to the notion of transnumeration where data visualization and statistical measures are used in service of making sense of a context to learn more about the context itself. In the context of variation different measures such as the range, interquartile range, and standard deviation can be used to represent variation as well as graphical displays such as boxplots, histograms, and bar graphs, which allow the investigator to “see” variation in the data. This relates back to statements earlier that the focus needs to be on statistics in service of reading and writing the world. Through visualizing variation in a data set one can begin to look for patterns and structures in the data that relay information about the context being measured. With the current explosion of technology in the world today, exploring data through visualizations and basic statistical measures has become increasingly easy to do even from the palm of your hand. This trend makes the notion of transnumeration increasingly important in statistics education and a crucial aspect of any vision of a critical statistics education as it is deeply intertwined in the process of statistical enquiry and relates to writing the world.

18.7.6 Problem Posing

A final aspect that is important for a critical statistics education is pedagogical and that is the practice of teaching through problem posing (Freire 1970). Problem posing does not mean simply giving students problems to solve. Instead, Freire (1970) describes it as a pedagogy where the teacher/student dichotomy is broken down, and both teachers and students collaborate in dialogue with one another. This pedagogy pairs well with teaching and learning statistics as teachers and students can explore issues together bringing their own prior experiences to bear to make sense of data based arguments. Furthermore, statistical investigations are based around problem posing as they begin with asking questions (Franklin et al. 2007). It is crucial that students be given opportunities to pose their own problems, which are meaningful and relevant to their lives that they can then investigate. Teachers in turn can bring their strong background in mathematics and statistics into the conversation to show different lens for making sense of issues and reading the world. More specifically in the case of fostering statistical literacy, students need experience posing such problems and then going through the stages of a statistical investigative cycle, ending with interpretations

of their results that include critiques of their process. Such experiences should include considering the implications the interpretations made may have in terms of the context and what actions might be suggested as a result.

18.8 Issues of Implementation

There are a number of significant issues related to implementing a critical statistics education perspective in school classrooms. One is the consideration of context. It is central to statistics and to critical mathematics education, yet the discussion of sociopolitical issues such as race, sexuality, or even political campaigns is generally considered very controversial in school classrooms and in some cases is forbidden by administrators or policy. Yet those are the kind of issues students face in their lives and that are visible in the world around them. Ignoring such issues within the walls of the mathematics classroom only perpetuates the paradox of relevance that Skovsmose (1994b) discussed. An important consideration here is what does this mean for the classroom teacher who is tasked to plan and carry out mathematics and statistics curriculum in the classroom?

To facilitate meaningful discussions around mathematical and statistical concepts and practices as well as their application to contextual issues is a significant task and requires some knowledge of the context being explored. This does not mean that teachers need to be experts in a multitude of content areas related to the contexts they explore to be able to carry on meaningful conversations and investigations with their students. However, they do need to have some familiarity with the major discussion points around such issues and the different viewpoints that are relevant. Furthermore, they should have taken some time to consider their own subjectivity towards such issues and how they might influence their instruction. Without prior consideration and reflection, classroom discussions could fall into ideological arguments with little basis in supporting evidence or more importantly in using mathematics and statistics to explore such issues. Considering contextual issues also means teachers need to be comfortable with not knowing everything that is being discussed in the classroom and to be comfortable in the role as learners along with their students and that students may at times challenge the teacher's own positioning in discussions. This relates directly to taking up a problem posing pedagogy (Freire 1970), discussed previously.

Another issue for implantation is related to teacher education and considering how to shape experiences for teachers to be prepared to take on the tasks necessary for implementing a vision of critical statistics education in their classrooms. Teachers are already expected to enter the classroom with content knowledge, and increasingly pedagogical content knowledge is being included in teacher training (Ball et al. 2008; Shulman 1987), but what is not common is discussion of what knowledge of contexts that can be investigated using mathematics and statistics might be useful or necessary.

It is not realistic to require teachers to have an extensive knowledge of contexts for the use of mathematics and statistics, particularly given the already overloaded curriculum of many teacher education programs. However, where advances towards such teaching could be made is in terms of pedagogy in methods courses, focusing on how to make sense of contexts using mathematics and statistics in the classroom and how to facilitate meaningful discussions around such issues. Furthermore, there is still a need for teachers to have more preparation in statistical concepts and practices. The *Statistical Education of Teachers* (Franklin et al. 2015) is taking strides in this direction. However, it will take many years for such policy to effect widespread changes in classrooms, and such policy documents do not explicitly advance a critical perspective of statistics education, which has been the focus of this chapter. What this means is that for teachers to be prepared and have the proper resources to facilitate experiences for students consistent with a critical statistical education, significant strides will need to be made at the teachers education level. A possible way to do this would be to work from the guidelines of the *Statistical Education of Teachers* (Franklin et al. 2015) providing pre-service teachers with statistics and statistics methods courses and focusing on the types of pedagogy discussed in this chapter. Furthermore, in statistics content courses an emphasize could be made on using statistics to investigate real real-world problems, which there has already been a call to do at the undergraduate level (Frankenstein 1994, 2009; Lesser 2007).

Taking up the type of pedagogy described in this chapter of course is easier said than done, and the reform movement in mathematics education has met serious challenges in a similar endeavor (Schoenfeld 2004). One such relevant issue is that of families, parents and their beliefs and values around what and how their children should be taught in schools. A number of mathematics reform movements in the U.S. have been crippled and collapsed by parents (Orrill 2016). This in part could be managed by not only working to create learning environments based on open dialogue between teachers and students but also by opening up such spaces through dialoguing with parents and the community as well. In a very real sense it is such open dialogue that is a hallmark of democracy, and to prepare students to be citizens in their democratic societies we can begin by creating learning environments that function as democracies where all vested parties have a say. Now this is not to say that such an opening up of learning spaces does not bring its own issues, but it does allow for dialogue and negotiation between parties to create a balance of perspectives and goals in shaping the learning environments for students to experience. That being said, taking this approach has not always worked in the past with reform efforts in mathematics, and there is a need for more research around how to initiate and sustain productive dialogue between all the stakeholders in education. Though I have only presented a few of the challenges I would argue they are some of the largest that need to initially be tackled to begin to shift towards both a critical mathematics and critical statistics education in schools.

18.9 Conclusion

This chapter presented the foundations and a possible vision for what a critical statistical education could be in K-12 mathematics classrooms. In moving forward however, there needs to be more work investigating what a critical statistics education could look like in the classroom. This type of work requires partnerships between statistics and mathematics educational researchers and mathematics teachers to develop a better understanding of how critical statistics education can be implemented in the classroom as well as what are some of the affordances and constraints of such implementation. Implications that this type of education has for mathematics teacher education also need to be considered and studied. For example, what type of experiences do mathematics teachers need to have to develop the flexible understanding of statistics and mathematics necessary for this type of teaching? Another question of concern is how to get important stakeholders of the mathematics community on board and involved with such changes.

Statistics' value in K-12 education with the goal of preparing students to become citizens in today's information based societies comes from the core practices that make up the statistical process: to pose questions, collect relevant data, analyze the data in the context of a problem (Franklin et al. 2007), and then verbalize the story that the data tell about an issue to others in a precise well principled argument (Abelson 1995). These practices situated in mathematics classrooms can begin to provide students with experiences in critically investigating and critiquing their own context in society, while developing the statistical concepts and practices that will enable them to make sense of their context. The task is dynamic and complex. The goal is not for students to completely understand or solve issues of society but to instead grapple with ideas using statistics to read and write the word and the world.

References

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: L. Erlbaum Associates.
- Apple, M. W. (1992). Do the standards go far enough? Power, policy, and practice in mathematics education. *Journal for Research in Mathematics Education*, 23(5), 412–431.
- Apple, M. W., & Beane, J. (2007). The case for democratic schools. In *Democratic schools: Lesson in powerful education* (2nd ed.). Portsmouth, NH: Heinemann.
- 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.
- Belham, M. (2016, May 16). Lies, damned lies, and Brexit statistics. *The Guardian*. Retrieved from <https://www.theguardian.com/news/datablog/2016/may/16/lies-damned-lies-and-brexit-statistics>. Accessed 5 May, 2017.
- Ben-Zvi, D., & Garfield, J. B. (2004). *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht; Boston: Kluwer Academic Publishers.
- Ben-Zvi, D., & Garfield, J. (2008). Introducing the emerging discipline of statistics education. *School Science and Mathematics*, 108(8), 355–361.

- Bergen, S. (2016). *Melding data with social justice in undergraduate statistics and data science courses*. Paper presented at the IASE Roundtable Conference 2016: *Promoting Understanding of Statistics about Society*, Berlin, Germany.
- Bishop, A. J. (1988). Mathematics education in its cultural context. *Educational Studies in Mathematics*, 19(2), 179–191.
- Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal for Research in Mathematics Education*, 33(4), 239–258.
- Boylan, M. (2016). Ethical dimensions of mathematics education. *Educational Studies in Mathematics*, 92(3), 395–409. <https://doi.org/10.1007/s10649-015-9678-z>.
- Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. New York: Routledge.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104(9), 801–823.
- Cohan, W. D. (2015, September). How Wall Street's bankers stayed out of jail. *The Atlantic*. Retrieved from <https://www.theatlantic.com/magazine/archive/2015/09/how-wall-streets-bankers-stayed-out-of-jail/399368/>. Accessed 5 May, 2017.
- Darder, A. (2014). *Freire and education*. New York, NY: Routledge.
- Davidian, M., & Louis, T. A. (2012). Why statistics? *Science*, 336(6077), 12. <https://doi.org/10.1126/science.1218685>.
- Davis, P., & Hersh, R. (1981). *The mathematical experience*. London: Penguin Books.
- Eichler, A., & Zapata-Cardona, L. (2016). *Empirical research in statistics education*. Cham: Springer International Publishing. Retrieved from <http://link.springer.com/10.1007/978-3-319-38968-4>.
- Engel, J. (2016). *Statistics education and monitoring progress towards civil rights*. Paper presented at the IASE Roundtable Conference 2016: *Promoting Understanding of Statistics about Society*, Berlin, Germany.
- Fienberg, S. E. (1992). A brief history of statistics in three and one-half chapters: A review essay. *Statistical Science*, 7(2), 208–225.
- Foucault, M. (1972). *The archaeology of knowledge and the discourse of language*. New York, NY: Pantheon Books.
- Frankenstein, M. (1994). Understanding the politics of mathematical knowledge as an integral part of becoming critically numerate. *Radical Statistics*, 56, 22–40.
- Frankenstein, M. (2009). Developing a critical mathematical numeracy through real real-life word problems. In L. Verschaffel, B. Greer, W. Van Dooran, & S. Mukhopadhyay (Eds.), *Words and worlds: Modeling verbal descriptions of situations* (pp. 111–130). Rotterdam: Sense Publishers.
- Franklin, C., Bargagliotti, A., Case, C., Kader, G., Scheaffer, R., & Spangler, D. (2015). *Statistical education of teachers*. American Statistical Association. Retrieved from <http://www.amstat.org/education/SET/SET.pdf>.
- 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.
- Freire, P. (1998). Education for critical consciousness. In A. M. Araujo & D. Macedo (Eds.), *The Paulo Freire Reader* (pp. 80–110). New York, NY: Cassell and Continuum.
- Freire, P. (1970). *Pedagogy of the oppressed*. New York, NY: Continuum.
- Freire, P., & Macedo, D. (1987). *Literacy: Reading the word and the world*. New York, NY: Taylor and Francis.
- Gal, I. (2002). Adults' statistical literacy: Meaning, components, responsibilities. *International Statistical Review*, 70(1), 1–52.
- Gattuso, L. (2006). Statistics and mathematics: Is it possible to create fruitful links. In *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador (Bahia), Brazil: International Association for Statistical Education and International Statistical Institute, Retrieved from http://iase-web.org/documents/papers/icots7/1C2_GATT.pdf.
- Gattuso, L., & Ottaviani, M. G. (2011). Complementing mathematical thinking and statistical thinking in school mathematics. 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. 121–132). New York, NY: Springer Science + Business Media.
- Giroux, H. A. (1989). *Schooling for democracy: Critical pedagogy in the modern age*. London: Routledge.
- Giroux, H. A. (1993). Literacy and the politics of difference. In C. Lankshear & P. McLaren (Eds.), *Critical literacy: Politics, praxis, and the postmodern* (pp. 367–377). Albany, NY: SUNY Press.
- Giroux, H. A. (2011). *On critical pedagogy*. USA: Bloomsbury Publishing.
- Gray, M. (2016). *Enhancing the understanding of statistics through student-conducted data gathering*. Paper presented at the IASE Roundtable Conference 2016: *Promoting Understanding of Statistics about Society*, Berlin, Germany.
- 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. (2013). Characterizing key developmental understandings and pedagogically powerful ideas within a statistical knowledge for teaching framework. *Mathematical Thinking and Learning*, 15(2), 121–145. <https://doi.org/10.1080/10986065.2013.770718>.
- Gutiérrez, R. (2013a). Why (urban) mathematics teachers need political knowledge. *Journal of Urban Mathematics Education*, 6(2). Retrieved from <http://ed-osprey.gsu.edu/ojs/index.php/JUME/article/download/223/148>. Accessed 6 May, 2017.
- Gutiérrez, R. (2013b). The sociopolitical turn in mathematics education. *Journal for Research in Mathematics Education*, 44(1), 37–68.
- Gutiérrez, R. (2009). Framing equity: Helping students “play the game” and “change the game”. *Teaching for Excellence and Equity in Mathematics*, 1(1), 5–7.
- Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, Latino school. *Journal for Research in Mathematics Education*, 34(1), 37–73.
- Gutstein, E. (2006). *Reading and writing the world with mathematics*. New York, NY: Routledge.
- Gutstein, E. (2013). Reflections on teaching and learning mathematics for social justice in urban schools. *Teaching Mathematics for Social Justice: Conversations with Educators* (pp. 63–78). Reston, VA: National Council of Teachers of Mathematics.
- Gutstein, E., Lipman, P., Hernandez, P., & de los Reyes, R. (1997). Culturally relevant mathematics teaching in a Mexican American context. *Journal for Research in Mathematics Education*, 28(6), 709–737. <https://doi.org/10.2307/749639>.
- Gutstein, E., & Peterson, B. (2013). *Rethinking mathematics: Teaching social justice by the numbers* (2nd ed.). Milwaukee, WI: Rethinking Schools, LDT.
- Harding, S. (1991). *Whose science? Whose knowledge? Thinking from women’s lives*. Ithaca, NY: Cornell University Press.
- Herbel-Eisenmann, B. A., Lubienski, S. T., & Id-Deen, L. (2006). Reconsidering the study of mathematics instructional practices: The importance of curricular context in understanding local and global teacher change. *Journal of Mathematics Teacher Education*, 9(4), 313–345.
- Katz, V. J. (2009). *A history of mathematics: An introduction*. Boston: Addison-Wesley.
- Labaree, D. F. (1997). Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal*, 34(1), 39–81.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465–491.
- Lesser, L. M. (2007). Critical values and transforming data: Teaching statistics with social justice. *Journal of Statistics Education*, 15(1), 1–21.
- Mathematics. (n.d.). In *Oxford English Dictionary online*. Retrieved November 27, 2017 from <https://en.oxforddictionaries.com/definition/mathematics>.
- Moore, D. S., & Cobb, G. W. (2000). Statistics and mathematics: Tension and cooperation. *The American Mathematical Monthly*, 107(7), 615–630.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- National Governors Association Center for Best Practices [NGA Center], & Council of Chief State School Officers [CCSSO]. (2010). *Common core state standards for mathematics*. Washington

- DC: Authors. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.
- Orrill, C. H. (2016). The process is just messy: A historical perspective on adoption of innovations. *The Mathematics Educator*, 25(2), 71–94.
- 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*. Monterrey, Mexico: International Commission on Mathematical Instruction and International Association for Statistical Education.
- Pfannkuch, M., Rubick, A., & Yoon, C. (2002). Statistical thinking and transnumeration. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.), *Mathematics Education in the South Pacific* (pp. 567–574). Auckland, New Zealand.
- Poling, L. L., & Naresh, N. (2014). Rethinking the intersection of statistics education and social justice. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education*. Flagstaff, AZ. Retrieved from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_C177_POLING.pdf.
- 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>.
- Schoenfeld, A. H. (2004). The math wars. *Educational Policy*, 18(1), 253–286.
- Shaughnessy, 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.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.
- Skovsmose, O. (1994a). Towards a critical mathematics education. *Educational Studies in Mathematics*, 27(1), 35–57.
- Skovsmose, O. (1994b). *Towards a philosophy of critical mathematics education*. New York, NY: Springer Science + Business Media.
- Skovsmose, O., & Valero, P. (2008). Democratic access to powerful mathematical ideas. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed., pp. 417–440). New York, NY: Routledge.
- Steen, L. (Ed.). (2001). *Mathematics and democracy: The case for quantitative literacy*. United States: The National Council on Education and the Disciplines.
- Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.
- Ullmann, P. (2016). *Communicating evidence? On the interaction of politics, data and the public*. Presented at the IASE Roundtable Conference 2016: *Promoting Understanding of Statistics about Society*, Berlin, Germany.
- 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*. Flagstaff, AZ. Retrieved from https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_PL1_USISKIN.pdf. Accessed 6 May, 2017.
- Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, 57(2), 74–79.
- Wager, A. A., & Stinson, D. W. (Eds.). (2012). *Teaching mathematics for social justice: Conversations with educators*. Reston, VA: National Council of Teachers of Mathematics.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–248.

Index

A

- Action research, 267, 268, 273
- Active learning, 8, 266
- Advanced Placement statistics, 145
- Affect, 15, 17, 18, 63, 102, 124, 131, 136, 245, 246, 256, 261, 313, 335
- Aggregate
 - aggregate aspects or features of a distribution, 73, 74, 76, 87, 202, 209
 - aggregate measures, 405
 - aggregate perspectives/view/point of view, 29, 71, 72, 76, 90, 118
 - aggregate reasoning, 71, 72, 74, 76, 77, 89, 90
 - aggregate reasoning with covariation (ARwC), 71, 72, 74, 76–78, 81, 83, 87–92
 - aggregate research question, 82
 - data aggregate, 72, 155
 - disaggregated data, 400, 401
- Algorithm, 67, 134, 228, 240, 376, 380, 403
- American Statistical Association (ASA), 124, 374
- Applet, 125, 127, 364
- Apprenticeship, 354
- Argumentation, 31, 35, 36, 48, 60, 61, 63, 64, 66, 266, 271, 322, 326
- Assessment
 - assessment task, 145, 367
 - formative assessment, 128, 290, 367
 - instructor assessment, 279, 280
 - self-assessment, 280, 354, 358
 - summative assessment, 274
- Association
 - bivariate association, 73

Attitude

- attitude toward/about mathematics, 204, 211
- attitude toward statistics, 332
- critical attitude, 315
- Attribute, 5, 13, 19, 77, 78, 80, 81, 83–86, 88, 90, 101, 102, 104, 106, 157, 275, 398, 403
- Availability, 261, 367
- Average, *see* center of a distribution

B

- Bayesian
 - Bayesian behavior, 5
 - Bayesian-type probability, 6, 20
 - Bayesian-type problem/task/situation, 3, 5, 6, 8, 18
- Belief
 - ability/self-ability belief, 339, 342, 343
 - conflicting beliefs, 341, 343
 - knowledge and beliefs, 56
 - personal belief, 63, 201
 - prior belief, 73
 - students' beliefs, 342
 - teachers' beliefs, 310, 313
- Big data, xi
- Big ideas, *see* concept
- Bootstrap, 401
- Boxplot, *see* graph

C

- Case study, *see* research method
- Cause and effect, 403
- CensusAtSchool, 187
- Center

- central tendency, 30, 47, 201, 228, 277, 316, 321, 376
- Center of a distribution
 - average, 30, 33, 47, 84, 89, 178
 - mean, 34, 52–54, 57, 62, 78, 83, 114
 - median, 34, 52–54, 57, 62, 78, 83, 114
 - mode, 53, 54, 61, 185, 202, 209–211, 219, 386
 - signal and noise, 74, 76, 80, 84–88, 91, 94, 101, 109–111, 119, 156, 186
- Central Limit Theorem (CLT), 267
- Chance
 - chance device, 100
 - chance experiment, 286, 289
 - chance-generating mechanism, 8, 10, 19
 - chance phenomenon, 3, 4, 20
 - chance variability/variation, 100, 102
 - data and chance, 98, 102, 113, 115, 118, 246
 - language of chance, 140
 - visualizing chance, 3
- Citizen Maths, 319, 352–357, 359–361, 363, 364, 366–370
- Clinical trial, 8
- Cognitive development, 127
- Collaboration, 254, 280
- College
 - college students/population, 181, 182, 246
 - teacher college, 204, 205, 221
- Common core standards, 396
- Common Online Data Analysis Platform (CODAP), 261
- Community
 - communities of practice, 266, 280
 - virtual community, 280
- Comparing groups
 - comparing data sets, 52
 - comparing distributions, 35, 128, 146, 202, 216, 246, 256, 271, 285
- Comparison
 - comparison (investigative) question, 180, 183–186, 188–194
 - comparison situation, 183, 186
 - group comparison, 53, 284–288, 290, 291, 295, 296, 298, 300, 302
- Computer
 - computer game, 104
 - computerized statistical model, xiii
 - computer lab, 276
 - computer simulation, *see* simulation
 - computer software, *see* software
 - portable computer, 334
 - tablet computer, 160, 162
- Concept
 - concept development, 27, 28, 34, 48
 - concept image, 124–126, 130, 131, 133, 140, 144, 148, 149
 - concepts-in-action, 28, 31, 34
 - key/pivotal/foundational statistical concept, 108, 125, 157, 261
 - mathematical concept, 4, 6, 28, 123, 148, 227, 395, 400, 401
 - probability (probabilistic) concept, 3, 20, 155
 - situative concept, 34
 - statistical concept, 7, 27, 28, 32, 34, 46, 77, 89, 102, 124, 125, 127, 146, 148, 153, 154, 158, 159, 162, 166, 167, 226, 227, 240, 246, 248, 257, 260, 277, 280, 315, 316, 323, 370, 373, 374, 392, 401, 406–408
- Conception
 - conceptions of statistics, 333
 - misconception, 3, 5, 8, 9, 13–15, 17, 19, 20, 73, 123, 124, 127, 128, 130, 133, 207, 225, 227, 229, 231, 233, 234, 237–241
- Conceptual
 - conceptual change, 107
 - conceptual development, 266, 370
 - conceptual framework, 56, 59, 100, 101
 - conceptual knowledge, 7, 181
 - conceptual mechanism, 204, 206, 208–210
 - conceptual structure, 74, 125, 126, 131, 132
 - conceptual understanding, 75, 124, 125, 168, 245–247, 257, 259–261, 310, 312, 341
 - levels of conceptual understanding in statistics (LOCUS), 250
- Conceptualization
 - conceptualization of measure, 27–30, 32, 46
- Conditional probability, 3–5, 9, 11, 17, 19, 273
- Confidence interval, 155, 156, 163, 164, 167, 168
- Connections (design and research project and learning environment), 76, 92
- Constructivism, 156
- Content analysis, 199, 205, 293, 294, 373, 379
- Context
 - real, real-life context, 141, 154, 315, 320, 373–378, 381–386
 - risk context, 51, 52, 57, 65, 67
 - statistical context, 148, 155, 236, 239, 241, 274

- Contextual
 - contextual knowledge, 27, 31, 82, 155
- Contingency table, 74
- Control, 20, 106, 108–110, 112, 116, 140, 148, 275, 357, 404
- Cooperative learning
 - collaborative learning, 370
 - collaborative learning environment, 342
- Correlation, 79, 85, 88, 90, 153, 160, 163, 164, 166, 167, 269, 270, 273, 275, 334
- Course
 - introductory statistics/probability course, 123, 144, 246, 253, 257, 329, 333, 334, 336
 - massive open online course (MOOC), 261, 351, 353, 358, 360, 364, 368–370
 - mathematics course, 9, 227, 272
 - online course, 351, 352, 356, 362
 - probability course, 17, 20, 273
 - statistics course, 13, 145, 246, 249, 257, 261, 272, 276, 278, 329, 334, 340, 343
 - undergraduate course, 249
 - university course, 312
- Covariation
 - reasoning with covariation, 72, 73, 90, 92
- Critical
 - critical citizen, 374, 375, 391, 392, 396
 - critical citizenship, 373–380, 383–386
 - critical education, 395
 - critical reflection, 126, 145, 146, 245, 246, 248, 250, 257–261
 - critical thinking, 359, 365, 374, 392, 394
- Cross-curricula statistics, 268
- Curriculum
 - curriculum guidelines, 311, 313, 325, 374
 - mathematics curriculum, 205, 271, 334, 391, 393–397, 399
 - national curriculum, xvi
 - standards, 271, 374, 394
 - statistics curriculum, 78, 174, 193, 221, 406
- D**
- Data
 - data collection, 9, 33, 78, 79, 103, 161, 205, 290, 311, 312, 314, 316, 326, 334, 335, 400
 - data creation, 204
 - data distribution, 75, 228, 233, 274, 334
 - data intensive linguistics, 265
 - data modeling, 391
 - data production, 404
 - data representation, 85, 103
 - data set, dataset, 51–54, 56–58, 61–63, 65–67, 73, 75, 91, 108, 124, 130–132, 181, 185, 202, 204, 212, 220, 222, 228, 247, 254, 255, 277, 283, 284, 287, 290–293, 301, 302, 316, 381, 403, 405
 - data structure, 71, 131, 405
 - real data, 10, 85, 104, 107, 117, 141, 284, 311, 312, 316, 320, 322, 325
 - realistic data, 140
 - simulating data, 288, 317
- Data analysis
 - data handling, 312, 315
 - exploratory data analysis (EDA), 99, 114, 115, 117, 154, 155, 226
- Data exploration
 - data investigation, 29, 77, 78, 100, 283, 286, 290, 311, 313, 323
- Decision making, 51, 56, 58, 61, 65, 153, 311, 391
- Density, 85, 88, 202, 236, 273, 288
- Design
 - design heuristic, 221, 222
 - designing investigation, 268
 - design of learning environment, 76, 91, 92, 98
 - design perspective, 145, 245, 248, 250, 260, 266, 278, 280, 309, 310, 312, 314, 325
 - design research, *see* research method
 - pedagogical design, 87, 91, 113–115, 266, 268, 280
 - task design, 35, 48, 76
- Design experiment, *see* research method
- Design research, *see* research method
- Deviation
 - mean absolute deviation, 132, 134, 136, 146, 226, 246, 249, 251, 253–260
- Didactical phenomenology, 33
- Discourse, 88, 114, 126, 145, 157, 245, 246, 248, 250, 253, 256, 258–261, 290, 309, 310, 313, 318, 319, 323–326, 392, 395, 403
- Dispersion, *see* variability
- Disposition
 - critical disposition, 374, 375, 392
- Distribution
 - center, *see* center of a distribution
 - distribution shape, 90, 130–132, 237, 240
 - distribution spread, 29, 74, 87, 90, 125, 130–132, 138, 202, 208, 211, 214, 217,

- 218, 221, 226, 229, 237, 241, 291, 295, 297, 310
 sampling distribution, 104, 105, 124, 143, 146, 248, 269–272, 274, 400, 401
- E**
 Education
 critical mathematical education, 391, 392, 394, 396, 401, 406, 407
 higher education, 226, 355
 mathematics (mathematical) education, 123, 126, 205, 260, 393, 394, 396, 400, 402, 407
 primary education, 205, 219
 probability education, 5
 statistics education, 6, 27, 29, 33, 72, 78, 103, 153–155, 168, 175, 200, 202, 271, 276, 283, 311, 330, 333, 374, 376, 377, 391–393, 396, 401–407
 teacher education, 200, 202, 219, 245, 261, 313, 326, 406, 408
 vocational education, 205, 206, 212, 213, 221, 368
 Epistemology/epistemological, 28, 33, 90, 158, 159, 162, 168, 169, 374
 Ethics/ethical, 9, 376, 394, 396
 Evidence, 6, 27–29, 31, 32, 35, 37, 46, 52, 54, 66, 72, 80, 83, 97, 99, 100, 155, 156, 159, 177–179, 185, 200–202, 205–208, 211, 214, 215, 218, 220, 221, 247, 250, 261, 279, 280, 314, 315, 321, 322, 360, 362, 366, 368, 404, 406
 Expectation, 14, 16, 57, 157, 164, 168, 179, 183, 248, 273, 276, 280, 324
 Experiment, 6, 33, 34, 131, 140, 141, 153, 159, 161, 163, 174, 177–182, 185–187, 231, 315, 317, 357, 400
 Experimentation, 127, 160
 Explanation, 13, 36, 42, 48, 60, 130, 154, 163, 167, 218, 220, 221, 235, 238, 239, 367, 368, 378
- F**
 Fathom, *see* software
 Focus group, 313, 365
 Framework, 7, 8, 10, 19, 27, 28, 31–34, 47, 87, 101, 127, 141, 145, 155, 157, 173, 174, 185, 188, 201, 204, 236, 247, 248, 250, 266, 269, 272, 273, 283–288, 290, 291, 293–295, 299, 301, 302, 310, 331, 375, 402
- Frequency, 4–7, 17, 34, 131, 140–142, 211, 212, 228, 229, 233, 237, 285, 294, 296, 297, 300, 384
- G**
 Gender, 18, 84, 124, 148, 185, 291, 295, 354, 355, 360–363, 377, 383, 395, 402, 403
 Generalization
 generalization beyond data, 154, 155, 160, 322
 generalization from data, 27, 153, 154
 generalization from sample to population, 155
 generalization process, 154
 probabilistic generalization, 27, 201, 209, 213
 uncertain generalization, 218–220
 Graph
 bar graph, 57, 128, 134, 229, 384, 405
 boxplot, 128, 291, 295, 300, 405
 case value graph, 255, 256
 dot plot, 128, 291, 295
 empirical probability graph, 11
 histogram, 57, 128, 229, 291, 405
 scatterplot, 74, 82, 127, 301
 Growing samples, 72, 76, 77, 80, 83, 87, 91, 104, 106, 108, 199, 200, 202, 203, 208–210, 219, 221, 222
 Guidelines for Assessment and Instruction in Statistics Education (GAISE), 124, 145, 193, 249, 269, 272, 273, 275, 276, 279, 280, 313, 402
- H**
 Heuristic, 104, 108, 199, 200, 202, 203, 219–222
 Hypothesis
 null hypothesis, 102
- I**
 Idea
 big (statistical) idea, 269–271
 central/core/foundational/key statistical idea, 67, 72, 76, 98–100, 108, 124, 126, 130, 290, 312, 316, 325
 powerful idea, 351–354, 359, 361, 362, 369, 400
 statistical idea, 56, 57, 78, 91, 99, 117, 148, 168, 193, 266, 269, 276–278, 290, 310, 313, 320, 321, 324, 378
 Inference

- statistical inference, 52, 71, 72, 97, 99, 100, 154–156, 228, 269, 270, 311, 312, 315, 397
- Inferentialism, 153, 154, 156, 158, 168
- Inferential reasoning, 200, 285, 320, 321
- Informal Inferential Reasoning (IIR), *see* reasoning
- Informal Statistical Inference (ISI), 27–29, 71, 72, 76, 77, 87, 97–101, 103–106, 113–116, 118, 153–155, 159, 163, 164, 166, 167, 169, 199–204, 206, 207, 213–215, 218, 219, 221, 222, 311, 312
- Inquiry
 - inquiry-based learning environment, 71, 77
 - statistical inquiry, 157, 265
 - statistical inquiry (enquiry) cycle, 175
- Interview, 7, 9, 17, 54, 61, 64, 66, 80, 159, 177, 231, 232, 234, 250, 259, 284, 285, 334–338, 341
- Investigation
 - data investigation, 29, 77, 78
 - experimental investigation, 202
 - growing samples investigation, 71, 77, 87
 - investigation cycle, 179, 193, 311
 - statistical investigation, 29, 76, 90, 104, 174, 175, 193, 266, 268, 273, 291, 311, 312, 315, 316, 330, 398, 404, 405
- K**
- Key statistical ideas, *see* concept
- Knowledge
 - content knowledge, 177, 200, 201, 230, 249, 260, 406
 - mathematical knowledge, 28, 227, 230, 231, 376
 - pedagogical content knowledge (PCK), 200, 230, 260, 261, 266, 406
 - pedagogical knowledge (PK), 268, 273
 - prior knowledge/existing knowledge, 107–109, 115, 148, 315, 342, 367
 - professional knowledge, 227, 230–232, 235, 241
 - statistical knowledge, 154, 185, 203, 225, 241, 268, 269, 277, 278, 330, 334, 351, 352, 354, 376–378, 384
 - teacher’ knowledge, 200, 201, 226, 227, 230, 231, 260, 310, 330
 - technological pedagogical content knowledge (TPCK), 266
 - technological pedagogical statistical knowledge (TPSK), 265, 266, 268, 269, 273, 276–280
- L**
- Language, 16, 28, 35, 54, 72, 81, 85, 87, 102, 131, 140, 146, 147, 155–158, 160, 162, 180, 183, 193, 201, 205, 209, 211, 213, 218, 219, 226, 236, 240, 312, 315, 316, 321, 325, 330, 335, 336, 375, 380, 395
- Learning
 - active learning, 8, 266
 - blended learning, 365
 - deep learning, 332
 - experiential learning, 126
 - learning goal, 32
 - learning mathematics, 148, 231, 330, 341, 352
 - learning statistics, 76, 123, 193, 279, 330–336, 338–343, 369, 402, 405
 - learning strategy/ies, 4, 332
 - learning theory, *see* theory
 - lifelong learning, 330, 342
 - virtual learning, 265
- Learning environment
 - blended learning environment, 266
 - cooperative/collaborative environment, 290, 310, 313, 326
 - dynamic learning environment, 315
 - inquiry/inquiry-based learning environment, 71, 77, 100
 - interactive dynamic environment, 125
 - statistical learning environment, 98, 311, 312
 - statistical reasoning learning environment (SRLE), 98, 99, 309, 310, 312, 314, 316, 319, 324–326
 - technology-enhanced learning environment, 98, 154
- Learning trajectory
 - actual learning trajectory, 79
 - hypothetical learning trajectory, 177, 179
 - statistical learning trajectory, 265
- Level
 - secondary level, 271
 - tertiary level, xii
- Literacy
 - adult literacy, 352, 353
 - critical literacy, 396, 402
 - data literacy, xvi
 - literacy perspective, 401
 - mathematical literacy, 395
 - quantitative literacy, 391, 392, 395
 - statistical literacy, 56, 250, 326, 329, 352, 353, 357, 359, 368, 370, 392, 401, 402, 405

M

Massive Open Online Course (MOOC), *see* course

Mathematical statistics, 249

Mathematics education, *see* education

Mean, *see* center of a distribution

Measure

contrasting measures, 35

conventional measure, 247, 276

formal measure, 75, 228, 241, 245–248

inadequate measure, 288

measure of center/central

tendency/centrality, 29, 132–134, 218, 221, 274, 325

measure of

variability/variation/spread/dispersion,

66, 136, 226, 228, 230, 237, 247, 254

nonstandard measure, 249, 251, 253, 259

shift model, 285, 288, 291, 295, 301, 303

situative measure, 31, 32, 35, 36, 38, 40, 41, 43, 46–48

statistical measure, 46, 100, 208, 228, 241, 311, 318, 321, 325, 385, 386, 405

summary measure, 128, 132, 138

Measurement

measurement error, 79, 160, 248

Misconception, *see* conception

Model

contrasting models, 48

covariation/covariational model, 82–85, 88, 90, 91

model for, 75, 84, 91, 104, 274, 309

model of, 53, 71, 75, 86, 88, 90, 91, 101, 103, 104, 126

model world, 102, 104–108, 114, 115, 117, 118

probability model, 8, 100

risk model, 67

statistical model, 75, 76, 100, 311, 320

theoretical model, 107

Modeling

integrated modeling approach (IMA), 97, 98, 100, 102, 103, 107, 113–115, 117, 118

statistical modeling, 275

stochastic modeling, 7

N

Narrative, 78, 127

National Council of Teachers of Mathematics (NCTM), 279, 394

National Science Foundation (NSF), 261, 268, 269, 280

Norm, 27, 153, 154, 157–159, 164, 166–168, 291, 301, 302

Normal curve, 230, 240

Normal distribution, 145, 216, 237, 240

O

Observation, 14, 32, 61, 62, 64, 65, 74, 103, 126, 127, 146, 147, 161, 181, 254, 261

Online course, *see* course

Outlier, 73, 75, 78, 83, 84, 86, 88, 90, 130, 146, 229, 249, 255, 256, 274, 285

P

Parameter, 4, 11, 12, 19, 155

Participation, 268, 335, 354, 356, 375, 396

Pattern, 27, 66, 72–75, 83, 85, 88, 100, 101, 109, 126, 131, 141, 146, 174, 204, 232, 246, 273, 300, 311, 352, 354, 369, 397, 402, 405

Pedagogical content knowledge, *see* knowledge

Pedagogy, 72, 87, 118, 148, 394, 405–407

Percentile, 364

Phenomenography, 333

Phenomenology, 33

Planning paradox, 357

Population, 10, 13–16, 18, 21, 71, 73, 75–77, 79, 80, 85–88, 91, 97, 98, 100–102, 104–106, 108, 117, 125, 126, 128, 130, 132, 133, 141, 143, 146, 155, 156, 179, 181–183, 185–193, 199–204, 206, 208, 209, 212, 214–216, 218–220, 222, 226, 228, 272, 291, 315, 316, 321, 352, 362, 364, 365, 368, 376, 383, 400, 401, 403

Pose/posing questions, 102, 130, 190, 408

PPDAC cycle, 175, 283

Practice

teaching practice, 202, 227, 309, 310, 312, 313, 322, 324–326, 374, 399

Practitioner, 7

Prediction, 6, 16, 35, 38, 72, 90, 91, 100, 141, 153, 155, 163, 168, 183, 202, 204, 205, 209, 211–215, 218–220, 311, 312, 315, 316, 378, 398

Probability

Bayesian-type probability, 6, 20
conditional probability, 3–5, 9, 11, 17, 19, 273

inverse probability, 5

joint probability, 12

mathematical probability, 4, 275

probability distribution, 102

probability language, 202

probability model/modeling, 8, 100
 probability of an outcome, 141
 probability of error, 99
 probability simulation, 102
 probability theory, 200, 205, 220, 221
 probability tree, 3, 4, 11, 18
 Purpose and utility, 76, 77, 88, 91, 155
p-value, 155, 191, 270, 277

Q

Qualitative, *see* research method
 Quantitative, *see* research method
 Quartile, 128, 133, 146, 295, 297, 298
 Question
 comparative investigative question, 174, 179, 185, 188, 192
 comparison (investigative) question, *see* comparison
 investigative/investigation question, 173–176, 178–183, 185–188, 192–194, 330
 multiple-choice question, 250, 367, 368
 open-ended question, 332–334, 338, 341
 research question, 4, 13, 32, 54, 76, 77, 99, 103, 159, 174, 178, 185, 188, 192, 203, 227, 268, 291, 293, 310, 315, 330, 336, 374, 379
 statistical question, 79, 173–175, 178, 193, 290, 311
 survey question, 104, 175, 183, 192, 335, 404
 teacher's question, 77, 108, 159, 161, 173, 174, 176, 177, 179–183, 185, 192, 193, 203, 204, 214, 220, 227, 234, 236, 237, 241, 250, 251, 253, 256, 257, 259–261, 292, 293, 310, 319, 323, 330, 405
 Questionnaire, 58, 79, 201, 290, 309, 319, 324, 325, 333, 335, 341, 344

R

Random
 random behavior, 130, 140, 141
 random data, 77
 random device, 8
 random event, 3, 20, 140
 random number, 18
 random outcome, 141
 random phenomenon, 8, 140
 random process, 102, 140
 random sample, 72, 75, 100, 102, 105, 109, 110, 112, 113, 117, 124, 126, 128, 130, 132, 141, 291, 311
 random sampling, 102, 104

random selection, 141
 random variable, *see* variable
 random variation, 4
 Randomization, 266, 274, 275, 277–279, 291
 Randomness, 4, 7, 19, 102, 125, 140, 141
 Reasoning
 aggregate reasoning, 71, 72, 74, 76, 77, 89, 90
 aggregate reasoning with covariation (ARwC), 71, 72, 74, 76–78, 81, 83, 87–92
 conceptual reasoning, 7
 determinist reasoning, 228
 formal reasoning, 75
 informal inferential reasoning (IIR), 27, 28, 72, 79, 80, 99, 100, 102, 118, 153–156, 158–169, 311, 312, 314, 315
 mathematical reasoning, 153, 158, 320
 probabilistic reasoning, 3, 5, 6, 18–20, 106, 228
 proportional reasoning, 5, 7
 reasoning level, 51, 60–63, 66, 67
 reasoning with data, 74, 77–81, 90
 reasoning with uncertainty, 98–100, 103, 106, 109, 112–114, 116
 statistical reasoning, 56, 71–75, 77, 97–99, 102, 104, 131, 145, 154, 155, 158, 193, 200, 232, 241, 245, 271, 278, 283, 284, 286, 287, 290, 291, 293, 295, 300, 302, 309–312, 314, 315, 318–321, 323–325, 378
 Reflection
 critical reflection, 126, 145, 146, 245, 246, 248, 250, 257–261
 reflective knowledge, 373, 376, 377, 381, 384–386
 Reform
 curricular reform, *xvi*
 educational reform, 407
 reform statistics education, *xi*, *xvi*
 Regression, 124, 269, 270, 334
 Representation. *See also* graph
 data representation, 85, 103
 graphical representation, 36, 57, 64, 146, 159, 202, 228, 230, 232, 234, 237, 239–241, 256, 311, 316, 322, 325, 382, 383
 informal representation, 321
 material representation, 181
 multiple representation, 124, 245, 248, 249, 253, 254, 256, 260, 274
 nonstandard representation, 260
 symbolic representation, 6, 10, 19, 148, 260

- Representation. *See also* graph (*cont.*)
 traditional representation, 260
 visual representation, 3, 6, 7, 11, 19, 126, 146, 241, 259, 260, 385
- Representativeness, 75, 78, 84, 155, 201
- Research method
 case study, xiii, 77, 88, 89, 91, 92, 102, 114, 116, 268
 classroom research, 302
 design experiment, 34, 36, 38, 39, 43, 177
 design research, xiii, 32, 33, 314
 qualitative, xiii, 103, 302, 319, 333
 qualitative content analysis, 293, 294
 quantitative, 282
 research methodology, 103, 177, 231, 334, 379
 teaching experiment, 159, 177, 178
 think aloud, 8, 9
- Risk, 3, 6, 51, 57, 61, 62, 64, 66, 67, 352, 357, 359
- S**
- Sample
 growing samples, *see* growing samples
 model-generated sample, 105
 random sample, 72, 75, 100, 102, 105–107, 109, 110, 112, 113, 117, 124, 126, 128, 130, 132, 141, 291, 311
 real sample, 71, 101, 102, 105, 117
 repeated sample, 98, 101, 107, 113, 115, 116, 220
 sample distribution, 54, 160, 201, 205, 208–211, 213–215, 218–221, 322
 sample size, 53, 72, 76, 101, 104–114, 116–118, 124, 126, 143, 146, 147, 199, 201, 202, 204, 205, 208–211, 214, 215, 217, 218, 220–222
 sample space, 273
 sample statistics, 100, 101, 124, 141, 143
 sample variability, 109, 201, 202, 217, 315
- Sampler (in TinkerPlots), 77, 80, 86, 91, 102, 104–106, 275
- Sampling
 biased sampling, 104
 empirical sampling, 274
 repeated sampling, 100, 105, 107, 109, 110, 115–117
 resampling, 275, 279
 sampling distribution, *see* distribution
 sampling method, 104, 201
 sampling space, 334
 sampling variability, *see* variability
- School
 elementary/primary school, 199–203, 220, 222, 283, 302
 high/secondary school, 145, 159, 187, 205, 221, 225, 227, 229, 231, 232, 240–242, 245, 246, 250, 253, 254, 259, 260, 266, 271, 273, 276, 278, 280, 283, 312, 334, 335, 353
 middle school, 33, 57, 67, 203, 250, 254, 260
 post-secondary, 392, 393
 private school, 57, 104
 tertiary, 246, 247
- Science education, *see* education
- Signal and noise, 74, 80, 85, 87, 88, 91, 101, 155
- Simulation, 8, 11, 12, 14–17, 19, 67, 100, 141, 143, 146, 266, 274, 315, 316, 354, 359, 366, 370, 401
- Situativity
 familiar situation, 364
 situative aspect, 31, 36, 41, 46, 47
 situative concept, *see* concept
 situative measure, *see* measure
 situative structure of phenomenon, 34
 situativity of knowledge, 28, 29, 34
- Social justice, 392, 395, 396, 399
- Software
 CODAP, 261
 CPMP-Tools, 268, 269, 273–275, 277
 dynamic interactive statistics software, 102
 Fathom, 254, 255, 268–270, 273–279, 285, 286, 288, 302
 Microsoft Excel, 334
 SPSS, 334
 statistical software package, 145
 TI Nspire, 145
 TinkerPlots, 35, 71, 72, 77, 79, 80, 82, 83, 85, 86, 91, 102, 104–107, 109, 115, 117, 249, 255, 268–270, 273–279, 283, 284, 286–293, 295, 297, 299–302, 311, 314–316, 318, 321–323, 325, 326
- Statistical inference, *see* inference
- Statistical inquiry, 157, 265
- Statistical knowledge. *See also* knowledge, 154, 155, 185, 203, 225, 241, 265–269, 277, 278, 330, 334, 351, 352, 354, 376–378, 384
- Statistical investigation, 29, 76, 90, 104, 174, 175, 179, 192, 193, 266, 268, 273, 291, 311, 312, 315, 316, 330, 398, 404, 405
- Statistically literate, 352, 392
- Statistical process, 265, 266, 312, 320, 384, 408

- Statistical reasoning, *see* reasoning
- Statistical Reasoning Learning Environment (SRLE), 98, 99, 309, 310, 312, 314, 316, 319, 324–326
- Statistical software packages, *see* software
- Statistical thinking, *see* thinking
- Statistics, 9, 29, 31, 34, 52, 56, 59, 60, 67, 75, 78, 98, 102, 107, 115, 125–128, 140, 143, 148, 149, 153, 155, 158, 159, 164, 175, 178, 181, 192, 193, 201, 202, 205, 211, 212, 219–222, 225–228, 230, 231, 234, 235, 240, 241, 245, 248–250, 256, 257, 259–261, 265, 266, 268, 269, 274–279, 283, 291, 297, 300, 301, 310–315, 320–324, 330, 331, 333–335, 337–343, 352, 353, 357, 364, 365, 370, 373, 374, 376–378, 386, 391–393, 396–403, 405–408
- Statistics education, *see* education
- Structure of Observed Learning Outcomes (SOLO) taxonomy, 53, 178, 189, 284
- Student
 - middle school student, 51, 55, 57
 - school student, 6, 52, 54, 200, 240, 245, 246
 - young learner/student, 71, 72, 98, 100, 102, 113, 116, 154, 155, 311
- Survey, 54, 100, 104, 175, 182, 225, 249, 268, 269, 275, 277, 278, 335, 341, 368
- Survey question, *see* question
- T**
- Table
 - data table, 161
 - frequency table, 5, 381
 - 2×2 /two-way table, 6, 7, 17
- Task
 - task design, 35, 48, 76
 - textbook task, 385
- Taxonomy, 178, 189, 284, 285
- Teacher. *See also* teacher education
 - elementary school teacher/primary teacher, 123, 145, 199, 201–203, 205, 219, 220, 222, 284
 - Indonesian pre-service EFL teacher, 329, 330, 340, 342, 343
 - in-service teacher, 159, 245, 265
 - instructor, 261, 279
 - middle school teacher, 250, 254, 260
 - pre-service teacher, 199, 200–204, 207, 210, 215, 218–222, 407
 - secondary teacher, 246, 247, 249, 265, 268, 269
 - teacher's knowledge, 200, 227, 230, 231, 260, 310, 330
 - teacher's perspectives, 245, 310, 313, 319, 324, 325
- Teacher education/statistics teacher education
 - initial teacher education, 205
 - online professional development
 - professional development, 241, 245, 246, 248–251, 256–261, 310, 313, 314, 326
 - professional growth
 - teacher preparation, 145, 261
 - teacher professional development cycle
- Teaching
 - teaching statistics, 177, 231, 242, 265, 273, 274, 276, 309, 313, 374, 376, 377, 392, 399, 400
- Technological knowledge, *see* knowledge
- Technology. *See also* software
 - graphing/graphic calculator, 124, 255, 274
 - interactive dynamic technology, 123, 124, 148
 - technological tools, 7, 19, 99, 103, 104, 107, 123, 271, 273, 274, 277, 290, 312, 313
- Textbook, 258, 373, 374, 378–380, 382, 384–386
- Theory
 - background theory, 153, 156, 168, 169
 - critical theory
 - domain-specific, 311
 - explanatory, 398
 - foreground theory, 278
 - framework, 8, 10, 19, 27, 250, 310, 331, 375
 - learning theory, 245, 248, 259
 - local instruction theory, 178
 - transformative learning theory, 245, 248, 259
- Think aloud, *see* research method
- Thinking
 - critical thinking, 359, 365, 374, 392, 394
 - mathematical thinking, 375, 376, 399
 - statistical thinking, 75, 128, 145, 174, 225, 226, 239, 241, 245, 266, 290–292, 302, 354, 378, 393, 399, 404
- Tinkerplots, *see* software
- U**
- Uncertainty
 - certainty, 57, 58, 140, 155, 201, 202, 204, 209, 210, 214, 216–219, 221, 226, 398
 - contextual uncertainty, 98, 101, 109, 110, 112–117

- statistical uncertainty, 98, 101, 102, 109, 110, 112–116
- Undergraduate statistics education, *see* education
- Understanding
 - conceptual understanding, 75, 124, 125, 168, 245–247, 257, 259–261, 310, 312, 341
 - deep understanding, 125, 155, 246, 248, 330
 - mathematical understanding, 246, 368
 - sense-making, 265, 280, 326
 - statistical understanding, 107, 245, 250, 261, 279
- V**
- Values on learning statistics, 329, 330, 332, 333, 337, 339, 341–343
- Variability
 - chance variability, *see* chance
 - dispersion, 228, 232, 236, 241, 316
 - interquartile range, 232, 246, 247, 249
 - measures of variability, dispersion, 90, 201, 226, 247
 - natural variability, 47
 - range, 28, 42, 52, 54, 55, 67, 81, 129, 146, 226, 228, 232, 237, 246
 - sampling variability, 72, 101, 102, 106, 108, 109, 113–116, 118, 130, 133, 201, 208, 210, 214–218, 220, 221, 400
 - standard deviation, 134, 135, 145, 226
 - variance, 226, 229, 246, 273
 - variation, xi, xiii, 51, 54, 56, 60, 99, 100, 127, 131, 142, 155, 158, 160, 240, 245, 248, 249, 261, 277, 284, 290, 310, 311, 312, 320, 322, 324, 325, 336, 343, 378, 398, 401
- Variable
 - lurking variable, 404
 - random variable, 273
- Video, 9, 34, 54, 148, 161, 177, 178, 180, 249, 257, 277, 280, 283, 284, 287, 289–291, 293, 295, 302, 357, 368, 370
- Virtual environment, 280
- Virtual module, 266, 267, 280
- Virtual simulation, 322
- Visualization
 - dynamic visualization, 4, 8, 18, 20