

Dani Ben-Zvi · Katie Makar *Editors*

The Teaching and Learning of Statistics

International Perspectives

 Springer

The Teaching and Learning of Statistics

Dani Ben-Zvi • Katie Makar
Editors

The Teaching and Learning of Statistics

International Perspectives

 Springer

Editors

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

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

ISBN 978-3-319-23469-4 ISBN 978-3-319-23470-0 (eBook)
DOI 10.1007/978-3-319-23470-0

Library of Congress Control Number: 2015958808

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2016

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.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

Contents

1 International Perspectives on the Teaching and Learning of Statistics.....	1
Dani Ben-Zvi and Katie Makar	
Part I Integrating the Contextual and the Statistical Worlds	
2 Young Students' Mental Modelling of Statistical Situations.....	13
Andreas Eichler and Markus Vogel	
3 Children's Wonder How to Wander Between Data and Context	25
Dani Ben-Zvi and Keren Aridor-Berger	
4 Supporting Students to Develop Concepts Underlying Sampling and to Shuttle Between Contextual and Statistical Spheres	37
Adri Dierdorp, Arthur Bakker, Jan van Maanen, and Harrie Eijkelhof	
Part II Learning Statistics at the School Level	
5 The Language of Shape	51
Pip Arnold and Maxine Pfannkuch	
6 Development of an Understanding of a Sampling Distribution.....	63
Eun-Sung Ko	
7 Korean High School Students' Understanding of the Concept of Correlation	71
A-Ra No, Soo-Yun Han, and Yun Joo Yoo	
8 Connections Between Statistical Thinking and Critical Thinking: A Case Study	83
Einav Aizikovitsh-Udi, Sebastian Kuntze, and David Clarke	

9	Tasks Associated to the Treatment of Tables at Elementary School and Its Level of Difficulty	95
	Soledad Estrella, Arturo Mena-Lorca, and Raimundo Olfos	
10	An Analysis of the Statistical Contents Covered In China, Singapore and Taiwan Mathematics Textbooks at the Primary Level	97
	Hak Ping Tam and OU Yung Chih	
11	Mathematical Modelling for Critical Statistics Education	99
	Luana Oliveira Sampaio and Maria Lucia Lorenzetti Wodewotzki	
12	Survey and Research on the Levels of High School Students' Critical Evaluation of Statistical Information and the Influence Factors	101
	Ma Ping	
13	The Conceptual Understanding of Variability in the Data Distributions	117
	Eun-Jeung Ji and Won Kyoung Kim	
14	Students' Misconceptions and Mistakes Related to Measurement in Statistical Investigation and Graphical Representation of Data	119
	Hyung Ju Yun, Eun-Sung Ko, and Yun Joo Yoo	
15	Overall Understanding of the Middle School Mathematics Course in Teaching Statistics Main Line	121
	Guan Jian, Bai Xue, Wu Peng, and Li Qiusheng	
16	Growth on Fourth-Grade Students' Mathematical Understanding of Average	123
	Jun Wu	
17	Exploring Dot Plot in the Perspective of Embodied Cognition	125
	Verônica Yumi Kataoka, Irene Cazorla, Eurivalda Santana, and Claudia Borim da Silva	
Part III Learning Statistics at the Tertiary Level		
18	Students' Difficulties in Understanding of Confidence Intervals	129
	Ana Henriques	
19	A Modeling and Simulation Approach to Informal Inference: Successes and Challenges	139
	Jennifer Noll, Mulugeta Gebresenbet, and Erin Demorest Glover	
20	Students' Understanding of Statistical Terms Having Lexical Ambiguity	151
	Soyeon Jung and Jihyun Hwang	

21 Student Understanding of Symbols in Introductory Statistics Courses 163
 Hyung Won Kim, Tim Fukawa-Connelly, and Samuel A. Cook

22 Two-Year College Mathematics Instructors’ Conceptions of Variation 175
 Monica Dabos

23 Students’ Sense-Making of Graphical Representation in a Basic Statistics Module..... 177
 Hui Teng Chia

Part IV Technology in Statistics Education

24 Bootstrapping Confidence Intervals..... 181
 Ross Parsonage, Maxine Pfannkuch, Chris J. Wild, and Kate Aloisio

25 Undergraduate Students’ Conceptions of Variability in a Dynamic Computer-Based Environment 193
 George Ekol and Nathalie Sinclair

26 Transforming Statistics Education Through ICT Application..... 205
 Tae Rim Lee

27 Selecting Technology to Promote Learning in an Online Introductory Statistics Course..... 215
 Megan Mocko

28 The Role of Technology in Indian Statistics Education: A Review 227
 D.S. Hooda and B.K. Hooda

29 Building Up the Box Plot as a Tool for Representing and Structuring Data Distributions: An Instructional Effort Using *Tinkerplots* and Evidence of Students’ Reasoning..... 235
 Luis Saldanha and Michael McAllister

Part V Statistics Teachers and Teaching

30 Integrated Reasoning About Statistical Variation: Secondary Teachers’ Development of Foundational Understandings 249
 Susan A. Peters and Ksenia S. Kopeikin

31 Preservice Teachers’ Difficulties with Statistical Writing..... 261
 Min-Sun Park, Mimi Park, Eun-Jung Lee, and Kyeong-Hwa Lee

32 Teachers’ Questions in the Statistics Class..... 271
 Lucia Zapata-Cardona and Pedro Rocha-Salamanca

33	Statistical Training of Pre-service Teachers with Application in School Practice	279
	Miriam Utsumi, Irene Cazorla, and Verônica Yumi Kataoka	
34	Statistics in Primary Education in Greece: How Ready Are Primary Teachers?	289
	Eugenia Koleza and Aristoula Kontogianni	
35	Developing Statistical Literacy (DSL): Student Learning and Teacher Education	299
	Hélia Oliveira, Ana Henriques, and João Pedro da Ponte	
36	Teacher Capacity as a Key Element of National Curriculum Reform in Statistical Thinking: A Comparative Study Between Australia and China	301
	Qinqiong Zhang and Max Stephens	
37	A Framework for Assessing Statistical Knowledge for Teaching Based on the Identification of Conceptions of Variability Held by Teachers	315
	Orlando González	
38	An Investigation into the Statistics Education of Preservice Mathematics Teachers in an Irish University	327
	Olivia Fitzmaurice, Ailish Hannigan, and Aisling Mary Leavy	
	Index	329

Chapter 1

International Perspectives on the Teaching and Learning of Statistics

Dani Ben-Zvi and Katie Makar

“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).

1.1 Introduction

Being able to provide sound evidence-based arguments and critically evaluate data-based claims are important skills that all citizens should have. It is not surprising therefore that the study of statistics worldwide at all educational levels is gaining more attention. The study of statistics provides students with tools, ideas and dispositions to react intelligently to information in the world around them. Reflecting this need to improve students' ability to think statistically, statistical literacy and reasoning are becoming part of the mainstream school and university curricula in many countries. As a consequence, statistics education is becoming a thriving field of research and curricular development. This book reflects this trend by introducing cutting edge research in statistics education done by scholars from twenty different countries across six continents. The studies in this book introduce and empirically tackle unique challenges—challenges that students face as they learn statistics, that teachers face as they teach students to understand and reason about data, and other important theoretical and techno-pedagogical challenges.

D. Ben-Zvi (✉)
Faculty of Education, The University of Haifa, Haifa, Israel
e-mail: dbenzvi@univ.haifa.ac.il

K. Makar
School of Education, The University of Queensland, St. Lucia, QLD, Australia
e-mail: k.makar@uq.edu.au

1.2 The Growing Importance of Statistics Worldwide

Statistics is vigorously gaining importance and recognition in today's society. Statistics is a central tool in moving science, economics, politics, schools, and universities forward. Quantitative information is omnipresent in media and in the everyday lives of citizens worldwide. Data are increasingly used to add credibility to advertisements, arguments, or personal and professional advice. Therefore, there is a growing public and policy consensus that being able to provide reliable and persuasive evidence-based arguments and critically evaluate data-based inferences are crucial skills that all citizens of the twenty-first century should have. All students consequently must become statistically literate (Gal, 2002) and be able to reason statistically—even at an informal level—as part of their compulsory and lifelong education (Watson, 2006). It is not surprising therefore, that attention has accelerated over the last decade to the development of data-based reasoning by the statistics, science, and mathematics education communities, as well as policy makers and the general public worldwide. Reflecting this essential need to improve students' ability to think statistically, statistical literacy and reasoning are becoming a necessary and important area of study that involves distinctive and powerful ways of thinking in nearly every field (Moore, 1998); however, the challenges of teaching and learning statistics are numerous.

1.3 The Challenges of Learning Statistics

Despite the increasing awareness of the importance of statistical literacy, statistics has been viewed by many students as difficult and unpleasant to learn. “Statistics anxiety” is experienced by as many as 80 % of graduate students (Onwuegbuzie, 2004). Many university instructors find statistics and research methods courses equally frustrating and unrewarding to teach. In schools, mathematics teachers often view statistics as a marginal strand in the mathematics curriculum and therefore minimize or ignore its teaching. As more and more students study statistics, teachers are faced with many challenges in helping these students succeed in learning and appreciating statistics. According to Ben-Zvi and Garfield (2004, p. 4), four key challenges dominate the field:

- Many statistical ideas and rules are complex, difficult, and/or counterintuitive. It is challenging to motivate students to engage in the hard work of learning statistics.
- Many students have difficulty with the underlying mathematics (such as fractions, decimals, algebraic formulas) and that interferes with learning the related statistical content.
- The context in many statistical problems may mislead the students, causing them to rely on their experiences and often faulty intuitions to produce an answer, rather than select an appropriate statistical procedure.

- Students equate statistics with mathematics and expect the focus to be on numbers, computations, formulas, and one right answer. They are uncomfortable with the messiness of data, the different possible interpretations based on different assumptions, and the extensive use of writing and communication skills.

Amid these challenges many teachers endeavor to enable students to develop statistical literacy, reasoning, and thinking. There appears to be a consensus among statistics educators and researchers that robust statistical literacy, reasoning, and thinking are the most important goals for students' learning statistics, and that these goals are not currently being fully achieved (see current review of research on teaching and learning statistics, Garfield & Ben-Zvi, 2007). The dissatisfaction with students' ability to think and reason statistically, even after formally studying statistics at the undergraduate and graduate level, has led to a reexamination of the field of statistics education.

As the statistics education discipline has evolved and become more distinct and separate from mathematics, changes have been called for in the teaching of statistics. The reform movement at the introductory statistics university level has long urged for instruction to be focused more on data and less on theory, and has called for changes in content (more data analysis, less probability), pedagogy (fewer lectures, more active learning), and technology (for data analysis and simulations) (Cobb, 1992; Gould, 2010; Moore, 1997). At the school level, there are increasing efforts worldwide to help students develop an understanding and familiarity with exploratory data analysis (EDA), modeling, and simulation rather than teaching them a set of separate skills and procedures. New pedagogical approaches are trialed, such as inquiry-based learning or project-based learning, combined with increasing efforts to integrate technology to reform learning (Biehler, Ben-Zvi, Bakker, & Makar, 2013). New K-12 curricular programs set ambitious goals for statistics education, including developing students' statistical reasoning and understanding (e.g., Australia—Australian Curriculum, Assessment & Reporting Authority, 2011; New Zealand—Ministry of Education, 2007; USA—National Council of Teachers of Mathematics, 2000; and Project 2061's Benchmarks for Science Literacy, American Association for the Advancement of Science, 1993).

Several factors have led to these current efforts to change the teaching of statistics at all educational levels. These factors include critical changes: new techniques of data exploration; changes in the use of technology and its growing availability at home and work; increased awareness of the implications of not developing students' ability to think or reason statistically, despite good performance in statistics courses; and concerns about the preparation of teachers of statistics at the K-12 and university level, many of whom have never studied applied statistics nor engaged in data analysis activities (Ben-Zvi & Garfield, 2004). However, despite reform efforts, many statistics courses at the university level still teach the same progression of content and emphasize the same development of skills and procedures. Many students still leave their course perceiving statistics as a set of tools and techniques that are soon forgotten. Even current methods of teaching continue to focus on the development of skills and have neglected to instill in their courses experiences that develop the ability to think statistically (Pfannkuch & Wild, 2004).

1.4 The Challenges of Teaching Statistics

The rising worldwide appreciation for the importance of an understanding of statistics, which means that the subject is now taught throughout school levels including primary school, has led to a paradigm shift in the conceptualization of statistics teaching and statistics teachers' professional development. Many teachers, whose experience is mainly grounded in descriptive statistics, are challenged by recent approaches and guidelines for statistics teaching and learning and new ambitious curricula programs for students. Such huge shifts in teaching approaches and thinking about the nature, role, and purpose of statistics require teacher educators to design and implement novel courses that will develop teachers' statistical thinking. It is also assumed that teachers need to acquire not only statistical knowledge but also professional knowledge to teach statistics, and that an effective and positive learning environment can develop in teachers a deep and meaningful understanding of statistics and ability to think and reason statistically (Pfannkuch & Ben-Zvi, 2011).

A recent book about this topic (Batanero, Burrill, & Reading, 2011) identifies new approaches to enhancing teachers' statistical literacy that bridge teacher education with teaching practice in the classroom. It advocates professional development and teacher education to attend to teachers' attitudes and beliefs (not just knowledge), strengthen teachers' statistical pedagogical content knowledge, and encourage different experiences and initiatives for teaching statistics, especially for primary teachers. For example, experiences which empower teachers to teach statistics by engaging them with real data and statistical investigations and connecting these experiences to their teaching practice and classrooms are critical to operationalizing these recommendations. Although some research directions are promising (e.g., Makar & Fielding-Wells, 2011), further study in statistics teachers' education is needed to extend and broaden the influence of such research.

1.5 The Challenges of Research in Statistics Education

Current trends and challenges of research in statistics education are represented in this book. Generally, we identify a tendency to follow the directions set by the reform movement of statistics education (Garfield & Ben-Zvi, 2008). This is evident in the emphasis on learning, teaching, and design of learning trajectories around the big ideas of statistics (Burrill & Biehler, 2011), challenging traditional rote forms of instruction and replacing them with active forms of learning which are typically enhanced by modern technology. More and more researchers are involved in evaluation of these interventions that aspire to create inquiry-based and collaborative learning environments for learners at all ages. These learning environments often incorporate authentic data (in many cases, real data, not merely realistic) that learners

collect themselves, organize and represent, analyze and infer, and communicate and argue findings and insights. The goals of developing students' statistical literacy, reasoning, and thinking are widely adopted, and more interestingly, interpreted and operationalized in a wide array of methods.

Many studies in this book explicitly study or rely on the use of technological tools not only to graph and calculate, but rather enhance creative learning, simulate abstract ideas, and experiment with models. From an affective aspect of learning, we see a broader range of attitudes toward statistics, including appreciation of the power of statistical processes, chance, randomness, and investigative rigor, and a propensity to become a critical evaluator of statistical claims. Finally, a range of alternative assessment methods are implemented or experimented to better understand and document student learning.

There are several current themes of great interest to researchers in the field, exemplified well in this collection of international studies. First, an ongoing empirical effort is taking place to understand the role of context in students' statistical reasoning and especially in their informal inferential reasoning (see the 2011 double Special Issue of *Mathematical Thinking and Learning* on this topic, Makar & Ben-Zvi, 2011). Research suggests that students typically attend either to their statistical knowledge or to their contextual knowledge, but find it challenging to integrate these knowledge bases (Makar, Bakker, & Ben-Zvi, 2011). Many of the studies in this book discuss these challenges of integrating the contextual world and the statistical world, evident in all stages of the data investigation: transforming the contextual research question into a data-based statistical question, developing appropriate measurement tools and data collection procedures, and transforming the data analysis results into meaningful and relevant insights on the context. These studies address in multiple ways the concerns about how statistical learners can be misguided by their contextual knowledge, pay attention to students' novel uncertainties related to the context rather than merely to statistical uncertainties, and suggest ways to support learners to better coordinate statistical and contextual knowledge.

Second, the integration of EDA and probability is freshly discussed in these studies. Fifty years ago, when statistics began to shift from being a course designed for a limited cohort of future scientists into its appearance in the broader academic curriculum, it still relied heavily on probability. The reinterpretation of statistics into separate practices comprising EDA and confirmatory data analysis (CDA, inferential statistics) freed certain kinds of data analysis from ties to probability-based models, so that the analysis of data could begin to acquire status as an independent intellectual activity (Tukey, 1980). The introduction of simple data tools, such as stem and leaf plots and boxplots, paved the way for students at all levels to analyze real data interactively without having to spend hours on the underlying theory, calculations, and complicated procedures. Computers and new pedagogies later completed the "data revolution" in statistics education (Ben-Zvi & Garfield, 2008). Studies in this book try to take advantage of new visualization, simulation, and modeling tools to reexamine the heart of statistics teaching and learning. It seems that the time is ripe for a new balance between approaches to EDA and probability,

perhaps through enhanced modeling and simulation approaches, but much more innovative design, experimentation, and field testing remains crucially needed.

Third, the number of studies about teachers in this book is impressive. It reflects the growing understanding that teachers are the driving force in reforming statistics education. No doubt, the key to prospering statistics education starts with teachers themselves. However, many teachers do not consider themselves well prepared to teach statistics nor face their students' difficulties (Batanero et al., 2011). The issues that are discussed in these studies are complex and varied, for example: teachers' statistical pedagogical content knowledge, teachers' attitude to statistics, the challenge of effective professional development, and classroom teaching practices (e.g., teachers' questions in class).

Fourth, technology continues to draw researchers' attention in growing numbers. Progress in the understandings of teaching and learning of statistical reasoning and the availability of innovative technological tools for learning statistics have enabled to integrate and readily capitalize on these advances (Biehler et al., 2013; Chance, Ben-Zvi, Garfield, & Medina, 2007). The studies on technology in this book provide a fresh perspective on advances in digital technologies and how they can support the development of students' statistical reasoning. Technology is used in these studies at the school level as well as in tertiary education, school-workplace vocational setting, public sector, assessment, and the development of teachers. Many challenges are still ahead of us in fulfilling the educational promise and added value of technologies to maximize the potential of learning technologies in statistics education.

Finally, but perhaps the diamond in the statistic education research crown, are the studies focusing on students' learning of statistics. We find it meaningful that the current studies emphasize new issues such as linguistic and discourse aspects of learning and understanding (such as ambiguity), application of knowledge out of school (e.g., border crossing in vocational education), and connecting critical thinking and statistical reasoning.

1.6 Looking Forward: The International Perspective

The impressive achievements of statistics education to improve its teaching and learning in a handful of countries have not yet arrived in all corners of the globe. Many countries still lack sufficient resources, updated curriculum materials, effective professional development of teachers, and current technologies, infrastructure essential to carry on the reform movement in statistics education. Citizens in these countries are especially in need of becoming literate consumers of data that are vital for improving the quality of life, monitoring and promoting social justice, economic growth and the environment. They are deserving, 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 the availability of high-quality technological tools for learning 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. An arena for collaboration in synergetic cross-country research and development projects can be created, and a sense of inclusiveness is nurtured and encouraged in this book. 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 six continents. It is a sound evidence for the growing numbers of enthusiastic and able scholars, the success of statistics education community worldwide, and the increasing recognition that statistics education is receiving in the educational world, especially in the mathematics and science education. 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 the international collaboration, sharing and contribution, to the mutual benefit of our future citizens of the world.

1.7 Introducing the Book

This growing interest in statistics education throughout the world is reflected in the research presented in this book. This collection was contributed by the members of Topic Study Group 12 (TSG-12) on Teaching and Learning Statistics at the Twelfth International Congress on Mathematics Education (ICME-12) that took place in July 2012 in Seoul. The researchers in TGS-12 came from 20 different countries and varied significantly by experience, background, and context. The enthusiasm and passion for statistics education and the sense of community among its members is reflected in this book. They all shared a common desire to improve statistics education by focusing their research on students' and teachers' conceptual understanding rather than rote learning and teaching, and developed collegial professional networks that supported the discussion of research in statistics education from around the world.

Talented, ambitious, and articulate young scholars and graduate students joined senior researchers from various countries in contributing to this book. It is encouraging to see the flourishing nature of statistics education research among young people, those who will soon be the future leaders of the statistics education community. Many of them are situated in remote areas or alone in their country, with

few opportunities to share and discuss their ideas, successes (and failures) with other researchers. Therefore, the TSG-12 meetings and the creation of this book were important opportunities for them to expose their work to others, get useful feedback, and start a dialog with other researchers.

The book is a valuable resource that presents the breadth and diversity of empirical and practical work done currently around the world. The diversity in the study themes is inspiring. A wide range of methods are used to respond to the research questions, from case studies of single students or teachers, aiming at deeply understanding their reasoning processes, to large-scale experimental studies attempting to generalize trends in the teaching and learning of statistics. Various epistemological stances are taken and used. Some researchers theoretically analyze the epistemology of a statistical concept, or web of concepts, to create a “knowledge map” of statistical ideas, which is then used to design innovative instructional materials or analyze teachers and students’ learning. Others refer to personal learners’ epistemology, the ways they view, apply, and control their own statistical knowledge.

All levels and types of education are studied, reflecting the growing spread of the teaching of statistics globally. The teaching and learning of statistics is portrayed in the book in designed settings for young children, students in formal schooling, tertiary level students, vocational schools, and teacher professional development. A diversity is evident also in the choices of what to teach (curriculum), when to teach (learning trajectory), how to teach (pedagogy), how to demonstrate evidence of learning (assessment), and what challenges teachers and students face when they solve statistical problems (reasoning and thinking). This large scope of diversity in research focus is fascinating and raises great hopes for the richness and growth of statistics education research in years to come.

The 38 chapters in this book are organized thematically in the following order:

1. Integrating the contextual and the statistical worlds
2. Challenges in learning statistics at the school level
3. Challenges in learning statistics at the tertiary level
4. Technology in statistics education
5. Statistics teachers and teaching

While most of the chapters are extended empirical reports, some are shorter reports (marked with an asterisk at the end of their title); these authentic and thought-provoking short reports are mostly authored from emerging scholars.

We hope that the integration of theories, empirical evidence, and instructional methods in these chapters can eventually help students and teachers to develop their statistical thinking worldwide. 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, perspectives, and key ideas of statistics. These are skills and knowledge that are crucial in the current age of information.

Acknowledgments We thank our collaborators Lisbeth Cordani (Brazil) and Arthur Bakker (Netherlands) for their support in reviewing and selecting the TSG-12 contributions. We also acknowledge TSG-12 local co-organizer Jangsun Paek and IPC Liaison Gail Burrill.

References

- American Association for the Advancement of Science. (1993). *Project 2061: Benchmarks for science literacy*. Washington, DC: Author.
- Australian Curriculum, Assessment and Reporting Authority. (2011). *Australian curriculum: Mathematics*. Canberra, Australia: Author.
- Batanero, C., Burrill, G., & Reading, C. (Eds.). (2011). *Teaching statistics in school mathematics: Challenges for teaching and teacher education*. New York: Springer.
- Ben-Zvi, D., & Garfield, J. (2004). *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht, The Netherlands: Kluwer.
- Ben-Zvi, D., & Garfield, J. (2008). Introducing the emerging discipline of statistics education. *School Science and Mathematics, 108*(8), 355–361.
- Biehler, R., Ben-Zvi, D., Bakker, A., & Maker, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, and F. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 643–690). Springer.
- Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero et al. (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 57–69). New York: Springer.
- Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1), 1–26.
- Cobb, G. W. (1992). Report of the joint ASA/MAA committee on undergraduate statistics. In *The American Statistical Association 1992 Proceedings of the Section on Statistical Education* (pp. 281–283). Alexandria, VA: American Statistical Association.
- Gal, I. (2002). Adults' statistical literacy: Meaning, components, responsibilities. *International Statistical Review, 70*(1), 1–25.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372–396.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.
- Gould, R. (2010). Statistics and the modern student. *International Statistical Review, 78*(2), 297–315.
- 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., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education*. New York: Springer.
- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, New Zealand: Learning Media.
- Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123–137.
- Moore, D. S. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association, 93*, 1253–1259.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

- Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment and Evaluation in Higher Education*, 29, 3–19.
- Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education*. New York: Springer.
- 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. 79–96). Dordrecht, Netherlands: Kluwer.
- Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician*, 34(4), 23–25.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.

Part I
Integrating the Contextual and the
Statistical Worlds

Chapter 2

Young Students' Mental Modelling of Statistical Situations

Andreas Eichler and Markus Vogel

2.1 Introduction

The review of Jones and Thornton (2005) referring to the recent decades of research in stochastics education yields the following results: The *Piagetian period* (e.g. Piaget & Inhelder, 1975) and the *Post-Piagetian Period* (e.g. Fischbein, 1975) generated empirical knowledge about the development of the probabilistic reasoning of students without schooling. Beyond that, in the *Contemporary Period* research provides empirical insights into the development of both probabilistic reasoning and statistical reasoning of students in the classroom as well as artificial learning environments deduced from research approaches according to intervention studies (Jones & Thornton, 2005). However, there is still a lack of empirical knowledge about the statistical reasoning of students without systematic schooling in situations of uncertainty that are not only focused on probability, and that are not mainly reduced to situations that could be modelled by Laplace experiments (cf. Mokros & Russel, 1995).

For this reason, our research approach aims for students' reasoning in situations of uncertainty that predominantly involve statistical data of random events and necessitate modelling with empirical (frequentist) probability. We are especially interested in how young students without schooling in stochastics deal with simple statistical situations of their daily life experience (Eichler & Vogel, 2011).

A. Eichler (✉)
University of Education Freiburg, Freiburg im Breisgau, Germany
e-mail: andreas.eichler@ph-freiburg.de

M. Vogel
University of Education Heidelberg, Heidelberg, Germany
e-mail: vogel@ph-heidelberg.de

Accordingly, essential questions of our research are: How do students deal with simple statistical situations of their workaday life? What features do they take into account when they build up a mental model to cope with the situation's demands?

We use an exploratory study to gain empirical knowledge concerning these research questions. Thus, the focus of our research is to develop our theoretical framework concerning reasoning of students not having been taught in stochastics, but not to prove systematically hypotheses that are deduced from an established theory. In this chapter, we first outline the foundation of our theoretical framework and discuss the theory-based development of tasks. Afterwards, we describe the method of the exploratory study and, finally, discuss exemplary results of our research.

2.2 Theoretical Framework

There are two crucial aspects determining the topic of our research questions: the statistical situation and the students, each with their individual characteristics. The relationship between available skills (including their development) of the students and the problem-determining conditions (including their mental cognition) is needed to be theoretically reflected on. Accordingly, our theoretical framework is based on two theories concerning the development of students' thinking (Siegler, 1996) and the structure of students' thinking according to the theory of mental models (Johnson-Laird, 1983). In the following two sections, we briefly explain both theories.

2.2.1 *Development of Thinking Processes*

In recent decades, scientists of different disciplines report research results contradicting the Piagetian stage development theory. For example, Biggs and Collis (1982, p. 21) state: "As we analysed the responses of hundreds of elementary, high school, and college students in several different subject areas, we found that the assumptions of stage theory [here the Piagetian stages of development are meant] did not hold".

Instead of a static categorisation of a student's thinking development, we are following Siegler (1996) whose research results demonstrate that children's thinking is far more variable than a staircase-based model suggests. He ascertained a lot of empirical evidence for children of a given age using a variety of strategies. This applies to different children as well as to an individual child (Siegler, 1996). Siegler described the development of children's thinking as "overlapping waves" with each wave corresponding to a different rule, strategy, theory, or way of thinking. Within this metaphor children's development of thinking is envisioned as a gradual ebbing and flowing of changing ways of thinking, with new approaches being added and old ones being eliminated (Siegler, 1996, p. 86).

Siegler (1996) aims for explaining the observed variability of children's thinking. This variability depends on different factors, like specific circumstances, requirements and available knowledge, which influence a child's concrete actions within different situations.

2.2.2 The Structure and Function of Thinking Processes: Mental Models

Highlighting the situation and its constituting characteristics when analysing children's thinking is also a main aspect of the theory of mental models (e.g. Johnson-Laird, 1983). Mental models are defined as representations of an entire situation in contrast to both the semantic representation of isolated propositions of a situation and the representation of superficial features of a situation (Kintsch, 1998). There is empirical evidence that mental models are not to be seen as fixed structures of memory being only recalled (Baguley & Payne, 1999). Following the information processing model of Schnotz and Bannert (1999), mental models are constructed individually according to a task and its requirements within a situation representing the structure as well as the function of the modelled object in an analogous way. Thus, mental models potentially show interpersonal differences.

Referring to the structure representing the static aspect of mental models, Johnson-Laird (1983, p. 156) states: "A mental model [...] plays a direct representational role since it is analogous to the structure of the corresponding state of affairs in the world—as we perceive or conceive it". An essential process of mental modelling a situation's structure is recognising the physical objects of the situation, e.g. a die and its characteristics, as well as the relationship of these objects and their characteristics. Given data are also to be seen as being part of a situation's structure because they represent results of a process having passed. Concerning the dynamic aspect of mental models, i.e. the function, Seel (2001) suggests that, when coping with demands of a specific situation, the learner constructs a mental model in order to simulate relevant aspects of the situation to be cognitively mastered. Thus, the function of mental models allows for deriving answers via mental simulation of systems by anticipating possible results given for example by throwing dice. Mental simulations do not result in quantitatively exact conclusions but in qualitative ideas about the expected outcomes of such simulations (De Kleer & Brown, 1983). These "qualitative simulations" (De Kleer & Brown, 1983, p. 155) require sense making about the system or process that should be simulated, its constituent components and their relationships.

According to Schnotz and Bannert (1999) mental models are hypothetical internal quasi-objects. Thus, they can only be inferred from observable information which represents mental modelling of a situation or task, the conditions of a students' specific situation (experience, pre-knowledge) and students' outcomes after working with tasks (written responses, videotapes).

2.3 Theory-Based Task Construction

According to our research focus and our theoretical approach we regard static and dynamic characteristics impacting on the structural relationships and dynamic behaviour of simple statistical situations when we thought about adequate tasks' construction. We distinguish:

Structure of a statistical situation: The tasks should refer to simple situations of decision making which could be expected easily to be understood by children of a given age. The structure is expected to be essentially judgeable by analyses of *objects* (including people) being involved in the situation as well as their relationship. In addition, given *data* representing former events should yield important information.

Function of a statistical situation: The tasks should demand for generating data (not necessarily in physical reality but in mind) and for *mental simulation* to come to a prognostic decision on basis of the available information. Thus, the function of a statistical situation is expected to be adequately estimated.

2.3.1 Model of Complexity Levels

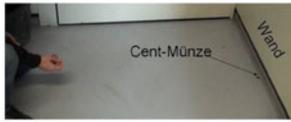
According to this distinction as well as to our theoretical framework we deduced theoretically a hierarchal model of complexity levels. The levels of task complexity in this model are developed according to the degree of the visibility of *objects* and *data* that mainly impact on the data generation in a given situation. Further, we understand an explicit demand for *mental simulation* in a task as further challenge for students. For the tasks we have used in our studies we identified four different levels (some tasks are explained exemplarily later).

From a theoretical point of view, it can be summarised: The more information about *data* and *objects* is given on the one side and the less demands with regard to mental simulation are requested, the more easily is a task to be estimated. Of course, this theoretical deduction has to be empirically proven. This was one of the goals of our empirical studies. Before we present some empirical results in the following section, we will exemplify the task construction by some concrete tasks.

2.3.2 Examples

We constructed four pairs of tasks. Each pair refers to a statistical situation which is well known for the students, e.g. in the money game task (Fig. 2.1). A pair of tasks is divided in two subtasks. One subtask refers to a small data sample (Fig. 2.1, left picture on right side). The other subtask refers to a bigger data sample (Fig. 2.1, right picture on right side).

In a game two players have to throw a coin ("Cent-Müne") near the wall ("Wand").



The first player throws the coin...



and the coin end's near the wall



The player, who has thrown the coin nearer to the wall, is the winner.

Andreas and Markus have trained to throw the coin and reached the following result



Andreas

Markus



Andreas

Markus

Who would be better in throwing a coin? Give a rationale for your answer:

Fig. 2.1 The money game task

Table 2.1 Criteria determining a model of the tasks' complexity (task names in brackets)

Four levels of complexity	Data	Objects	Mental simulation
Lowest (money game)	Given	Given	No explicit demand
Low (car)	Given	Given	explicitly Requested
High (M&M)	Given	Partly given	explicitly Requested
Highest (die)	Not given	Partly given	explicitly Requested

The tasks are from different levels of complexity with regard to the different criteria displayed in Table 2.1. For example, in the money game task (Fig. 2.1) both the data as well as the objects which mainly impact on the situation are given. Thus, the process and results of data generation are made visible. In the money game task, there is no need for mental simulation in terms of mentally anticipating results. However, the formulation "Who is better (at throwing a coin)?" demands for a decision which contains at least indirectly a mental simulation based on the situation's structure comprising objects and data. Summarising with regard to the model of the tasks' complexity (Table 2.1), the money game task is on the lowest level: data and objects and, thus, the data generation are visible. Moreover, the task does not include an explicit request for mental simulation, i.e. a prediction of a future event.

Examples for more complex levels were the so-called car task (low level) and the die task with the highest level of complexity according to the model shown in Table 2.1. With regard to be short of pages these tasks are presented in a nutshell only giving an overview (for details and for the other tasks see Eichler & Vogel, 2012).

2.4 Method

Our research approach and our main research questions mentioned earlier are part of a research programme which includes both quantitative as well as qualitative studies. In the following, we will especially focus on those analyses of one study which we have carried out for getting answers with regard to the tasks' construction and accordingly to the students' performances in dealing with these tasks.

2.4.1 Participants

In the study we report on, 134 students of grade 4 (primary school, about 9–10 years old) and grade 6 (secondary school, about 11–12 years old) participated. In Germany, there are different types of secondary schools. Thus, the sample for grade 6 is divided into students from secondary schools with potentially lower performances (Hauptschule; coded with HS) and schools with potentially higher performances (Gymnasium; coded with Gym). All students participating in this study have not been taught in stochastics before.

2.4.2 Material and Procedure

All students were asked to solve a set of four tasks. These four tasks were from different levels in terms of our model. On each level a pair of tasks is divided in two subtasks comparably to the example of the money game task mentioned earlier. Overall, there were two different task sets with each task consisting of one of the two subtasks mentioned earlier. In each class half of the students' group get one of the two sets of tasks. We randomised the students' assignment to one of the two sets of tasks. The students had 45 min to work on the tasks. By each task they were asked for a decision or an answer and for a rationale. We avoided any intervention in terms of giving students further information about the tasks.

2.4.3 Data

Concerning the students' performances in dealing with the tasks mentioned earlier the frequency of different levels of the quality of students' rationales was of primary interest. Thus, these rationales were coded for using an adaption of the SOLO model of Biggs and Collis (1982) including the following complexity levels of responses (in comparison to a specific adaption of Watson & Moritz, 2000): prestructural level, unistructural level, multistructural level and relational level. We illustrate each level by one student's rationale concerning the money game task (see Table 2.2, we assign the category 0 if the students did not provide any rationale). The coding was done independently by two researchers to confirm reliability.

This first step of the analysis is aimed to arrange the students' solutions for each task in a hierarchical structure to facilitate investigating differences in these solutions regarding age. In a second step, we interpreted the students' rationales differentiated by our SOLO adaption levels to identify potentially different types of mental models underlying the students' solutions and to characterise these types (for details see Eichler & Vogel, 2012).

Table 2.2 Codes and examples to the codes

Code	Example
Code 1, prestructural level: no or, respectively, no relevant justification of the response	“Markus is better” (grade 6, HS)
Code 2, unistructural level: justification of the response contains a single feature of the situation's structure concerning objects or data	“Markus is better, because three coins are close to the wall” (grade 6, HS)
Code 3, multistructural level: justification of the response contains more than one, although isolated features of the situation's structure concerning objects or data	“Markus because he has some coins close to the wall. However, he has also a lot of coins far away” (grade 6, HS)
Code 4, relational level: justification of the response contains interrelated features of the situation's structure concerning objects or data	“Andreas is better. On average, he has the most coins at a point close to the wall. By contrast, Markus has his [coins] close to the wall and more far away [from the wall]”

2.5 Results

2.5.1 *Quantitative Results for Proving Aspects of the Theoretical Framework*

We have reason for assuming that the complexity level of the tasks representation, which is defined by the visibility of objects and data in the task and the request of mental simulation, correlates with the students' performance.

Figure 2.2 illustrates some computations of the comparison of the students' solution for each pair of tasks. These results fit our theoretical model of the tasks' complexity. Thus, a task of a theoretically defined lower level of complexity seems to be also empirically seen more difficult for the students: The students' performances decrease. This effect proves to be significant particularly if more than one step of increased complexity is considered. When the comparison over two steps of the complexity levels is regarded, all differences prove to be statistically significant as well as to be (adequate) practically significant (Cohen's $d \approx 0.5$). Figure 2.2 (right side) shows the p -value of a t -test referring to the means that represent the students' performance of two different tasks.

Although the one-step differences between the low and high level as well as between high and highest level are also remarkable they proved not to be statistically significant. Deeper analyses of students' rationales lead to the assumption that the car task (low level, see Table 2.3) sometimes was interpreted in an unintended way, which might hinder statistical significance. Eliminating the car task from the analysis the aforementioned differences are without exception significant.

With regard to age-related differences we found that although students of grade 6 achieve slightly higher than students of grade 4 (Fig. 2.3, left), these differences are statistically not significant (t -test; $p = 0.227$). This finding meets the expectations according to the situational paradigm of our theoretical framework, or said it the other way round: If we had found statistically significant differences, we have had empirical reason for reconsidering the theoretical foundation.

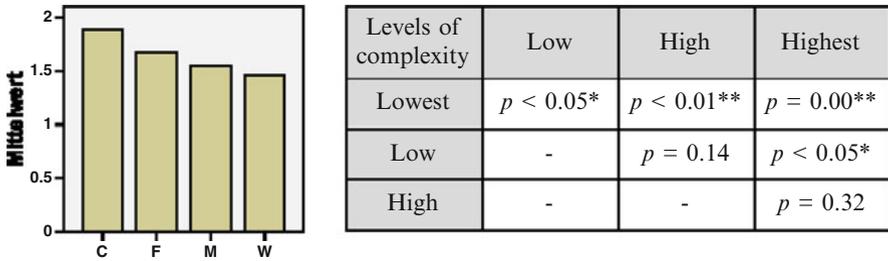


Fig. 2.2 Comparison of the tasks concerning the representation complexity

Table 2.3 Examples of tasks

Task	Car task (low level)	Die task (highest level)
Situation		
Question	A player starts an automatically accelerated car several times beginning from a fixed starting line. Each time the cars' end position is marked on the floor. Mark the position(s) where the next car (the next two cars) could stop. Give a rationale for your answer	In this game, the player wins when he gets a three in the next throwing of a die. A cuboid die and the ordinary die are available. Which of the dice would you take for the next throw? Give a rationale for your answer
Judgement	Data: given; objects: given (player, car); mental simulation: requested	Data: not given; objects: partly given (dice); mental simulation: requested

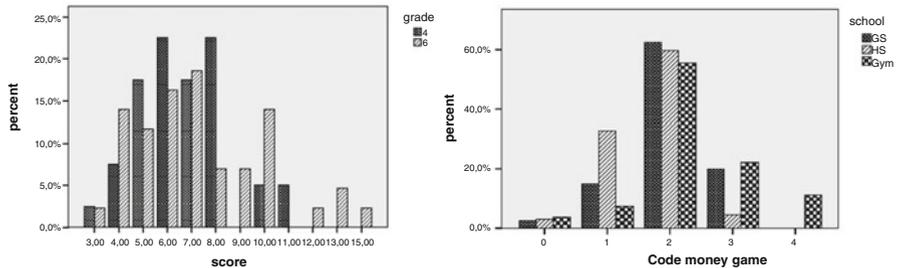


Fig. 2.3 Differences in students' performances

Besides from that we also looked for differences with regard to school form to enrich database for further studies. Here, we found significant differences on all complexity levels except from the low level including the car task (lowest level: $F=8.688, p<0.001$; low level: $F=1.355, p=0.262$; high level: $F=4.540, p=0.012$; highest level: $F=5.851, p=0.004$; distribution of, e.g. the money game task, see Fig. 2.3, right).

2.5.2 *Qualitative Results to Develop the Theoretical Framework*

Besides the results being presented to prove aspects of our theoretical framework quantitatively, we gained insights into students' reasoning from qualitative analyses. We briefly discuss most important results for purposes of illustrating possible directions of theory development.

The students' rationales differ with respect to their quality coded according to our adaption of the SOLO model. However, the students' rationales also differ within a specific model level, in particular concerning the students' perception of *objects* and *data*.

In all tasks the students identify different objects (human or non-human), as well as different characteristics of an object to mainly impact the data generation, but the students' also differ in their identification within a level of the SOLO model. For example, concerning the die task about 25 % of the students' identify the dice's rolling to mainly impact on the result of throwing the dice whereas also 25 % of the students identify the characteristics of the dice's sides:

Student GS27: "I would take the die on the right side [the ordinary die], because it is a square" (prestructural level).

Student Gym27: "The die on the right [the ordinary die], because it rolls better" (prestructural level).

Just as well, the students differ in their perception of a given data distribution. For example, referring to the money game task, 10 % use the minimum of the coin distance for their written rationales, and about 45 % use one of each, the centre and the maximum of coin distances. The differences are existent also in one level of the SOLO model:

Student GS16: "Markus is better, because he has two coins [directly] located at the wall" (unistructural level).

Student GHS24: "Andreas has performed better, since he is closer [to the wall] if the rear coins are included" (unistructural level).

To analyse differences in the students' perception of objects and data in a given situation with uncertainty seems to be crucial for two reasons: On the one side, the analysis of the students' perception of both objects and data yields an insight of students' reasoning beyond their categorising according to the SOLO model. On the other side, this analysis could potentially provide a theory based and empirically guided development of a diagnostic tool. Thus, the better a student is able to adequately perceive the main characteristics of objects and data in a given situation the more adequate seems to be his mental model and, in particular, his mental simulation of a situation. However, we found that the students' perception of a statistical situation is strongly influenced by the situation itself. For example, we have proved different levels of the tasks' complexity. However, the tasks' complexity seems to explain the students' reasoning as a group only, but not a student as an individual.

For example, about 50 % of the students provide an increase of the quality of their rationale from one task to another although the complexity of the first task is lower than that of the other task.

2.6 Discussion and Conclusion

In this report, we explained our research into young students' mental modelling of statistical situations by embedding our research questions in the field of existing research into stochastic education. We identified the characteristics of a statistical situation and a student's reasoning as being constitutive for mental modelling within a specific situation of uncertainty. Accordingly, our theoretical framework is based on theories of development of thinking process of Siegler (1996) and mental model theory (e.g. Johnson-Laird, 1983) because these theories allow for highlighting the dependence of reasoning of a specific situation. Based on this, we deduced theoretically a hierarchal model of complexity levels of tasks representing (simple) statistical situations and developed two sets of tasks which exemplify this model.

Summarising the quantitative analyses of our study we can state with regard to the hierarchical model that the theory-based construction has been empirically proven in general but in detail some modifications have to be reflected on when revising material for further studies. With regard to the students' performances we found no remarkable differences between fourth and sixth graders but with regard to school forms. These findings we read in being consonant with implications of our theoretical framework.

The results of our qualitative analyses concerning this study affirm the quantitative results and yield interesting more in-depth information about students' mental modelling: Analysing the students' rationales qualitatively in detail allowed us to distinguish different qualities of argumentations on the same level of our SOLO adaption with regard to the problem situation being resolved into objects, data and mental simulation as determining elements. Thus, according to the situational paradigm of our theoretical framework the claim for individual and problem-specific considerations could be met in a more adequate way.

This gives us reason to be confident for being able to develop a theoretically founded and empirically proven diagnostic tool in our future work. Furthermore, our database leads us to focus in further studies on questions of long-term effects and classroom implementation when we ask for improving young students' mental modelling within statistical situations of their workaday life.

References

- Baguley, T. S., & Payne, S. J. (1999). Memory for spatial descriptions: A test of the episodic construction trace hypothesis. *Memory and Cognition*, 27, 962–973.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic.

- De Kleer, J., & Brown, J. S. (1983). Assumptions and ambiguities in mechanistic mental models. In D. Gentner & A. Collins (Eds.), *Mental models* (pp. 155–190). Hillsdale, NJ: Erlbaum.
- Eichler, A., & Vogel, M. (2011). Mental models of basic stochastic concepts. *Proceedings of Seventh Congress of the European Society for Research in Mathematics Education*. Rzeszów, Poland: University of Rzeszów. <http://www.erm.eunito.it/>.
- Eichler, A., & Vogel, M. (2012). Basic modelling of uncertainty—young students' mental models. *ZDM*, 44(7), 841–854.
- Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht, The Netherlands: Reidel.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge, England: Cambridge University Press.
- Jones, G. A., & Thornton, C. A. (2005). An overview of research into the teaching and learning of probability. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 65–92). New York: Springer.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press.
- Mokros, J., & Russel, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26(1), 20–39.
- Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. London: Routledge & Kegan.
- Schnotz, W., & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen. *Zeitschrift für Experimentelle Psychologie*, 46(3), 217–236.
- Seel, N. M. (2001). Epistemology, situated cognition, and mental models: “Like a bridge over troubled water”. *Instructional Science*, 29(4–5), 403–427.
- Siegler, R. S. (1996). *Emerging minds. The process of change in children's thinking*. New York: Oxford University Press.
- Watson, J. M., & Moritz, J. B. (2000). The development of concepts on average. *Focus on Learning Problems in Mathematics*, 21, 15–39.

Chapter 3

Children's Wonder How to Wander Between Data and Context

Dani Ben-Zvi and Keren Aridor-Berger

3.1 Introduction

Aiming to make statistics meaningful and motivated for students, statistics teaching increasingly becomes context based, using authentic activities that are significant for the students (Garfield & Ben-Zvi, 2008; Wild, Pfannkuch, Regan, & Horton, 2011). One result of this trend is the growing interest in studying the role of context in developing students' statistical reasoning (Makar & Ben-Zvi, 2011). This chapter discusses the relations between data and context in learning to make informal statistical inferences (ISIs) by examining a pair of fifth-grade students' emergent understanding in a growing samples approach. We briefly review the literature on data and context, Informal Statistical Inference (ISI), Informal Inferential Reasoning (IIR), and the growing samples approach. We present several episodes that demonstrate students' transitions between the context and data worlds, and their growing understanding of the ways to integrate and use the two worlds.

3.2 Literature Review Data and Context

Statistics is the science of learning from data (Moore, 2005) that involves data production, data analysis, and statistical inference (Moore, 1997). Exploratory Data Analysis (EDA) was suggested by Tukey (1977) to make sense of data by organizing, describing, representing, and analyzing them—relying on informal analysis, visual displays, and technology (Cobb & Moore, 1997; Garfield & Ben-Zvi, 2008). In EDA we search iteratively for patterns and trends to gain more insight about the data while combining local and global views of them (Ben-Zvi & Arcavi, 2001;

D. Ben-Zvi (✉) • K. Aridor-Berger
Faculty of Education, The University of Haifa, Haifa, Israel
e-mail: dbenzvi@univ.haifa.ac.il

Konold & Higgins, 2003). The purpose of a statistical investigation is to exploit data to gain insights about a realistic event (Makar & Ben-Zvi, 2011) in its own context world (Wild & Pfannkuch, 1999). Thus, the recognition of data as “numbers in context” (Moore, 1990, p. 96) is critical to the development of statistical reasoning (Langrall, Nisbet, Mooney, & Jansem, 2011).

The notion of context has been studied in various disciplines and was given different meanings (e.g., Boaler, 1993). Our interpretation of the *context world* in the current study relies mainly on the idea of *context knowledge*—the knowledge about the context of the situation from which the statistical investigation arises (Gal, 2002). Important part of the context knowledge is the knowledge about the process of data collection (Pfannkuch, 2011).

The integration between the preacquired knowledge emanating from the context of the problem (such as beliefs, theories, dispositions) and the knowledge arising from the data allows interpreting and explaining the data rather than just manipulating it (Moore, 1990). Moreover, the frequent tension between these two types of knowledge can elicit the development of new insights about the problem in its context (Dierdorff, Bakker, Eijkelhof, & van Maanen, 2011; Makar, Bakker, & Ben-Zvi, 2011).

A statistical investigation can be described as a tapestry, in which its warp threads are the statistical knowledge, and the woof threads are the context knowledge (Cobb, 1999).

Statistical reasoning that does not fully consider these crisscross knowledge bases is not complete. Experienced researchers make ongoing transitions between the data world and context world during the statistical investigation. This transfer from data to context helps explain phenomena seen in the data in its context, while the opposite transfer helps answer new questions about the data and change the researcher’s initial understanding (Wild & Pfannkuch, 1999). Konold and Higgins (2003) describe this process as “a give-and-take conversation” between the hunches of the expert researcher about the investigated problem and the “story” that the data tells about these hunches.

These analogies emphasize the importance of the context in a statistical investigation, suggesting that context that is integrated in authentic environment and driven by authentic goals can scaffold the development of IIR by providing the investigators with a common language, through which they can better express their statistical ideas. However, the use of context might be an obstacle when, for example, the application of the context investigation as the main evidence allows the investigator to ignore the data (Makar et al., 2011).

3.3 ISI and IIR

A statistical inference is a generalization of the data in hand expressed by probabilistic language and evidenced by and extends beyond the data (Makar & Rubin, 2009). *Informal Statistical Inference* (ISI) is a theoretical and pedagogical approach

for developing statistical reasoning, while connecting between key statistical ideas and informal aspects of statistical inference by making such inferences informally (Garfield & Ben-Zvi, 2008). The main goal of teaching ISI is to deepen the understanding of the purpose and the gain that can be driven of the data and its interpretations (Makar & Rubin, 2009).

The reasoning process leading to making ISIs is called *Informal Inferential Reasoning* (IIR, Makar et al., 2011). IIR is a cognitive activity engaged in formulating generalizations (e.g., conclusions, predictions) from random samples of data using various statistical tools, while paying attention to evidence and uncertainty (Ben-Zvi, Bakker, & Makar, 2015). The development of students' IIR can bridge informal data exploration to formal statistical inference later (Ben-Zvi, Gil, & Apel, 2007). IIR can be nurtured by an inquiry-based learning environment with suitable tasks, tools, teacher scaffolds, and inquiry drivers (such as doubt, explanation, and resolution of cognitive conflicts) that focus on statistical concepts, statistical tools, and context knowledge (Makar et al., 2011). An example for a useful educational approach that can support the development of IIR is the “growing samples” task design.

3.4 Growing Samples

A key element of the current design of the educational materials is the approach of *growing samples*—an instructional idea mentioned by Konold and Pollatsek (2002), worked out by Bakker (2004, 2007), and elaborated by Ben-Zvi (2006). Starting with small datasets, students go through investigations of random samples with increasing sizes, taken from the same population (e.g., $n = 10$ cases, a class $n = 30$, a grade $n = 90$, a school $n \approx 500$), where they are asked to make *informal inferences* for each of them and to *predict* about their conclusions for an even larger sample. The goal for students is to develop their IIR by experiencing the strength and constraints of different size samples, and by examining the roles of context and data and their connections. This approach can be helpful in supporting coherent reasoning with key statistical concepts such as data, distribution, variability, tendency, and sampling (Bakker, 2004). It can also help students observe aggregate features of distributions, identify signals in them, account for the uncertainty in their inferences, and provide persuasive data-based arguments (Ben-Zvi, 2006).

3.5 Research Question

The goal of this case study is to characterize fifth-grade students' evolving understandings and use of the data and context worlds and the relationship between them, while they engage in drawing ISIs in growing samples investigations. In particular, we focus on illuminating the ways in which the connections between the two worlds

evolve and are expressed. Given this context, the two main research questions are: (1) How does the process of building connections between the context and the data worlds look like among young students? (2) How does this process develop over time?

3.6 Method

To address these questions we draw on data from the 2010 *Connections* Project in three Grade 5 classrooms in Israel. In this Project (2005–present) a group of researchers and teachers design and study an inquiry-based learning environment to encourage the development of statistical reasoning in grades 4–6 using *TinkerPlots* (Konold & Miller, 2005).

3.7 The Setting and Participants

The fifth-grade students participated in five extended data investigations (each lasting 2–3 lessons of 90 min). In each investigation, students iteratively posed research questions, organized their sample data using *TinkerPlots*, and interpreted it in order to draw informal inferences. Each lesson included a short whole class preview discussion about the investigated topic, an extended hands-on data investigation supported by *TinkerPlots* in small groups, and a closing whole class presentation and discussion of students' informal inferences. The students investigated data collected by themselves in a survey of four grades in their school (270 cases, 33 attributes, such as free time activities, body dimensions).

According to the growing samples design, students were given updated larger sample and asked to compare it to the previous sample and draw inferences informally. These informal inferences were derived by “what-if” questions in order to speculate what can be inferred about the next and larger sample taken from the same population. Students started investigating a sample of eight students from their class, continued to 27 cases (the entire class), then 81 cases (a grade level), and finally 270 cases (four grades). We focus on one pair of academically successful and articulate boys—Liron and Shay—following the development process of their statistical reasoning about data and context. Their investigations were fully videotaped using *Camtasia*TM to capture both their computer screen and discussions.

3.8 The Episodes

The episodes were selected from Shay and Liron's first two independent data investigations with *TinkerPlots*. In the first episode, the pair studied issues of free time (e.g., what students do in their free time) using a sample of eight students from their

class (including themselves) with 11 attributes. They were guided by a handout that included questions about sampling and inference, e.g., “Would the conclusions you have reached apply also to a larger group of students in your class, for example, half of the class? Please explain.” In the second episode, the sample size was increased to 27 (their whole class) and they were asked to investigate if their conclusions still held for the larger sample.

3.9 Data Analysis

The videos were carefully observed, transcribed, translated from Hebrew to English, and annotated for further analysis of the development of students' reasoning about data and context. We used interpretive microgenetic method (Siegler, 2006) taking into account verbal, gestural, and symbolic actions within the situations in which they occurred. Interpretations were discussed until consensus was reached.

3.10 Results

The following results illustrate the boys' statistical reasoning about data and context while making informal inferences, from initial primary reliance on data or context (Episodes 1 and 2) to a growing level of integration of data and context (Episodes 3 and 4).

3.10.1 Episode 1: Focusing Primarily on Data

In the first activity the boys analyzed a small sample of eight cases. Frustrated from their initial inability to make any sensible inference from the small sample, Shay and Liron searched for an attribute in the dataset whose graph was not too “spread out,” as Shay phrased it. Only when they looked at a plot of the number of after-school activities per week (Fig. 3.1), they became more confident about drawing a conclusion from the data.

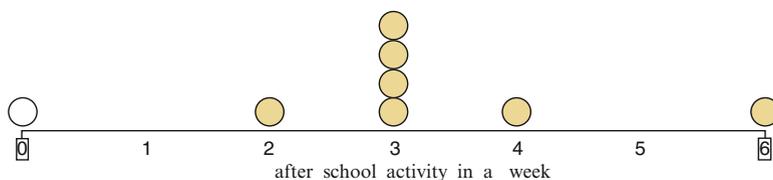


Fig. 3.1 A TinkerPlots stacked dot plot of the number of after school activities per week

34	S	Hmm, we got that it's usually three, that in a week there are three kids ^a
35	L	Wait a minute, there is one [child] ^b with six, you know. I bet you it's Shira [a female student in their class]
36	S	How do you know?
37	L	It doesn't matter. Four, four [three] after-school activities
38	S	Well, in short ...
93	L	Three is the biggest. It is the most frequent
40	S	According to what we see, the most frequent here is three. What we see is that it can be said also that the average here is three. The average here is three

^aThe translation from Hebrew to English was made to preserve the authenticity of the original utterances at the expense of correct English phrasing. Differences between Hebrew and English connotations of words were extensively discussed

^bInsertions in square brackets represent our best guesses of what students mean given the context

Shay and Liron are encouraged by the clear signal they identified in the after-school activity graph (Fig. 3.1). However, their response to this finding was different. Liron was aware of the problem context when he focused on one student [35] and the mode [37]. Unlike him, Shay viewed the data as numbers by focusing on the mode and average with no reference to the attribute in question or the context. He seems to examine the graph as a mathematical object detached from its context, and a few minutes later, when he saved this file, he even used a wrong title not related to the context. We regard these initial negotiations with the role of context in the unknown EDA field as an aspect of an enculturation process (e.g., Ben-Zvi & Arcavi, 2001): entering and picking up the points of view of a new discipline.

3.10.2 Episode 2: Focusing Primarily on the Context World

After several investigation cycles on different attributes, Shay and Liron felt they have extracted all available information from the data and moved on to report their work in writing. When they were requested to formulate an interesting question on free time based on the data at hand, Shay suggested: "How much free time do you have? I don't know how to formulate this. I mean those things you have to go to, like school and after-school activities, which you have them, and have to go to them at a certain time." In formulating this question, Shay ignored the data at hand (that did not include such an attribute) and reasoned with his context knowledge. He also offered ways to collect new data in order to answer it. When the interviewer pointed that they were to use their given data, Shay complained that the given sample "doesn't help," nevertheless the boys tried formulating another question.

103	S	So I think of [another question], I think I have an idea: Do you feel ...
104	L	That you have free time?
105	S	Yes. That you have free time
106	L	No, it is not [such a good idea] ...

107	S	No, I think that it is [a good idea]. If you find out there are kids that feel they are stressed ... Wow, there are so many kids who feel stressed. This is not a good thing—you need to give them freedom. Or if you find out that there are kids that feel free, then you’ll see, then you’ll see, wow, yes! [Writing down this question, “do you feel that you have free time?”]
108	S	[Reading from the handout] OK, “Organizing and analyzing the data.” But we don’t have data!

In spite of the interviewer’s explicit request to constrain themselves to data in the sample, Shay suggested a question about an attribute that did not exist in this sample, based on his context knowledge and personal experience. His abductive explanation [107] (Gil & Ben-Zvi, 2011) eventually convinced Liron to accept this proposal. It seems that their disappointment of what the “data world” (represented by the sample) offered them so far [108] led them to focus their reasoning primarily on the context world.

3.10.3 Episode 3: Context and Data Worlds Becoming Partly Connected

In the second data analysis activity, the sample was increased to 27 cases (the whole class). The students were asked to examine whether their previous inferences still held for the larger sample. The following dialog took place when they compared the graphs of the two samples (Fig. 3.2).

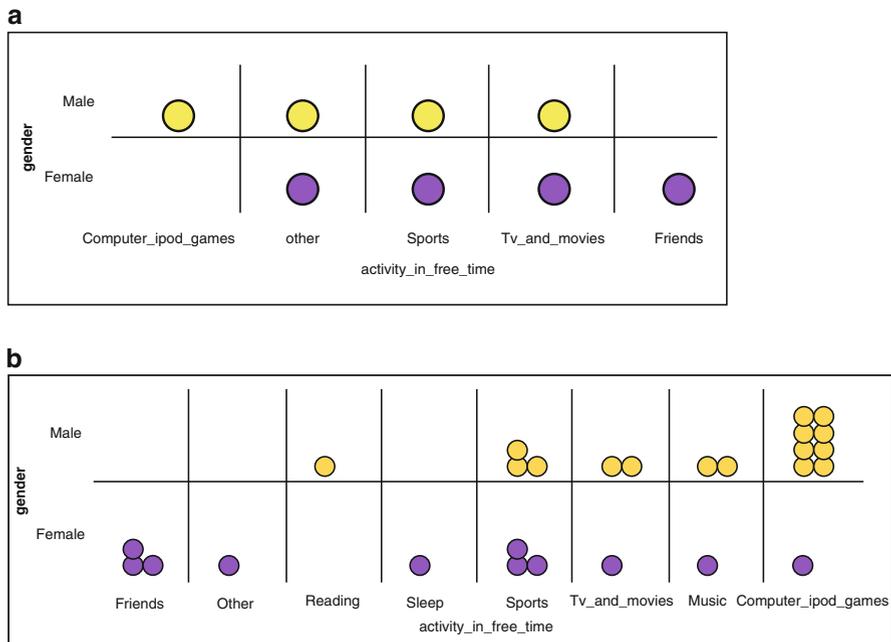


Fig. 3.2 Activity in free time by gender: (a) $n = 8$, (b) $n = 27$

359	S	Now, you see [looking at Fig. 3.2b]?... I see my hypotheses materializing! Boys like computers more than girls do. Girls like going to friends more than boys do. Where is sports?
360	L	Sports they are both together [Boys and girls like sports equally]
361	S	Wow, that's it, right
362	L	You are right, it is as if boys are into...
363	S	I was sure that boys are into sports more than girls
364	L	Yes, me too
365	S	But it [looking at Fig. 3.2b] is exactly the same [as Fig. 3.2a]! It is similar to the previous data [the small sample], look!
366	L	Yes, and still, right, because this is 27 kids from our same class

The beginning of this dialog derived its power mainly from the data and its graphical representations. Shay was astonished by the similarity between the plots of the two samples (Fig. 3.2) and the reinforcement it provided to their previous hypotheses and inferences [359]. This contradicted his previous anticipation that the small sample cannot tell him anything about the larger sample. However, one aspect of the data was disturbing for them: girls like sports equally to boys [360]. Here the data conflicts with their contextual preconception that, “boys are into sports more than girls” [363]. A choice between the data and their context knowledge had to be made. In accepting that boys and girls like sports equally, Shay's reasoning seems to rely more on the data than on his personal context knowledge. But this was not always the situation. For example, the following dialog took place when Shay and Liron addressed similar questions concerning the whole school population.

605	Int.	Now I'm interested to hear your description of the whole school... who starts? Liron
606	L	[The whole school is] very different [than our class sample]
607	Int.	Very different? Why? Different in what?
608	L	Different in what? Let's say the kids of the school are from grades 7–9; they do more sports during the breaks. I see them all the time playing soccer and tag. All sorts of stuff like these [pointing with the cursor at the sports category in the plot, Fig. 3.2b]
611	S	I think, I want to say: I agree with Liron that if we define a larger range of ages, then there will be also [more] boys, because boys like sports more than girls
612	Int.	Really? Do you claim that what seems here the same [boys and girls in sports] will be [different when the sample is increased]?
613	S	[Ignores the question] And also there will be boys who like to go to friends, a lot, like how it is [pointing with his hand toward the screen]
614	int.	Which here [the current sample, Fig. 3.2b] there is none [no boy likes to go to friends]
615	S	No one

When asked to describe the distribution of the whole school, Shay and Liron had to imagine a larger population with a larger range of ages (grades 1–9). Once again they paid attention to the conflict between the data and their contextual preconception that boys prefer sports more than girls. This dialog, unlike the previous one, derived its power mainly from their context knowledge about children in different ages [608, 611]. In the context of a prediction task [605], when data about other age levels were not available, their reasoning was dominated by their context knowledge, while keeping an eye on the sample data by frequent pointing at its plot and emphasizing the difference between the distributions.

3.10.4 Episode 4: Integrating Context and Data

So far, we presented Shay and Liron's shuttling back and forth between the data and context worlds. Toward the end of the second investigation, the boys integrated the worlds more often, by providing abductive explanations, raising conjectures or drawing informal inferences, in which data and context were both taken into account. For example, the following dialog took place when the boys addressed a prediction question concerning the fifth-grade population. Liron speculated that the whole grade's graph would maintain the proportion among the categories as in the class' graph except for minor changes.

537	Int.	Try to think of something typical [in the distribution of the whole grade]
538	L	The computer. These are modern times; most of the kids prefer computers over music
539	S	Or over reading books
540	Int.	And is this the picture you see here [in the class, Fig. 3.2b] too?
541	L	Yes
542	Int.	You have just said something general. You have said: "most of the kids"
543	L	Computer games, trrr tik [pretends counting the cases in the computer category, Fig. 3.2b], and music is three, it's a lot
544	S	Computer games are the most
545	L	And reading is just me, only I did this activity

The interviewer's question at the beginning of the dialog referred to the data world [535]. However, Liron integrated both context and data in his answer. Beginning from his context knowledge, he explained why children preferred a computer over watching television or reading a book [538–9]. He used the data at hand as evidence [543] to support his claim, comparing the frequencies of attributes and his personal outlying case as the only student that preferred reading.

3.11 Discussion

The examples brought above from Shay and Liron's investigations provide some insights on young students' building an understanding of the roles and relations of the context and the data worlds in an inquiry-based learning environment. Based on interpretive microgenetic analysis of video data, we sketched a nonlinear back and forth reasoning process of the students, strewn with many difficulties and challenges. The episodes presented the students' progress from initial transitions between the two worlds to their gradual meaningful integration. In our case, students' reasoning derived initially more often from their context knowledge rather than from the data at hand. During this long and complex stage, the students oscillated between the two worlds, putting aside or completely rejecting one of the worlds and returning to it later. We speculate that this behavior can be attributed to the small sample size used at the first activity (according to the growing samples design) and students' lack of fluency in graphicacy and in EDA. As the data analysis activities progressed and different investigations were experienced, students made more and more use of the connections between the worlds. These emergent connections were present in their explanations, predictions, and informal inferences, in which both worlds were considered.

This development in students' reasoning seems to be supported by the growing samples curricular design, the interviewer's interventions, and the social interactions between the students. The growing samples design helped increasing students' confidence in their ability to infer from the data and explain their informal inferences by weaving the two worlds. The use of prediction questions especially supported students' reasoning since they had to imagine what the population would look like based on both the data at hand and their context knowledge. They were thus forced to use both worlds simultaneously. Situations of conflict between the data and context were particularly catalyzing, in a similar manner to what is described in other studies (e.g., Ben-Zvi, Aridor, Makar, & Bakker, 2012). This process was enhanced also by the interviewer's persistence on asking the students' for clarifications of their prediction by providing data-based evidence and abductive explanations. Students' collaborative work also assisted in this process especially in situations in which each of the boys inferred from a different perspective. This elicited fruitful discussions that contributed to the development of their reasoning about data and context.

This concise description of the way young children wander between the data and context worlds while learning to make ISIs is far from being exhaustive and therefore still full of wonder. In our ongoing research, we intend to cultivate the above ideas and questions in order to shed more light on the role of different types of contexts, instructional situations, and learning environment designs in helping students confidently integrate data and context in informal inferential reasoning.

References

- 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. (2007). Diagrammatic reasoning and hypostatic abstraction in statistics education. *Semiotica*, 164, 9–29.
- Ben-Zvi, D. (2006, July 2–7). Scaffolding students' informal inference and argumentation. In A. Rossman, & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching of Statistics (CD-ROM)*, Salvador, Bahia, Brazil. 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, 35–65.
- Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM*, 44(7), 913–925.
- 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, August). 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 (SRTL-5)*. University of Warwick, Coventry, England.
- Boaler, J. (1993). The role of contexts in the mathematics classroom: Do they make mathematics more “real”? *For the Learning of Mathematics*, 13(2), 12–17.
- Cobb, G. (1999). Discussion of “let's use CQI in our statistics programs”. *The American Statistician*, 53(1), 16–21.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly*, 104, 801–823.
- Dierdorff, A., Bakker, A., Eijkelhof, H. M. C., & van Maanen, J. A. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning*, 13(1), 132–151.
- Gal, I. (2002). Adults' statistical literacy: Meaning, components, responsibilities. *International Statistical Review*, 70(1), 1–25.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.
- Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning*, 13(1), 87–108.
- Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.
- Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration (version 1.0) [Computer software]*. Emeryville, CA: Key Curriculum Press.
- Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33(4), 259–289.
- Langrall, C. W., Nisbet, S., Mooney, E. S., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning*, 13(1), 47–67.
- Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13(1), 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.

- 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.
- Moore, D. S. (1997). *Statistics: Concepts and controversies*. San Francisco: Freeman.
- Moore, D. S. (2005). Learning from data. In R. Peck, G. Casella, G. Cobb, R. Hoerl, D. Nolan, R. Starbuck, & H. Stern (Eds.), *Statistics: A guide to the unknown* (4th ed.). Belmont, CA: Thomson.
- Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, 13(1), 27–46.
- Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon & R. M. Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Cognition, perception, and language* (6th ed., vol. 2, pp. 464–510). Hoboken, NJ: Wiley.
- Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–265.
- Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference (with Discussion). *Journal of the Royal Statistical Society, Series A*, 174, 247–295.

Chapter 4

Supporting Students to Develop Concepts Underlying Sampling and to Shuttle Between Contextual and Statistical Spheres

Adri Dierdorp, Arthur Bakker, Jan van Maanen, and Harrie Eijkelhof

4.1 Shuttling Between Contextual and Statistical Spheres

This chapter deals with a fundamental challenge in educational design: to stimulate students to shuttle back and forth between a contextually phrased problem and statistics. The importance of moving between the contextual world of life and the statistical world of symbols is stressed in many different areas. Many contextual problems can be solved with the help of statistics. However, the results of computation or modeling need to be evaluated on their merits and validity in the context. More generally, Wild and Pfannkuch (1999, p. 228) stressed the significance of “shuttling between the contextual and statistical spheres.”

From the literature, we know that such shuttling is not easily promoted in students. Ainley, Pratt, and Hansen (2006) propose to focus on the purpose of tasks and utility of what is learned. A purposeful task is one that has a meaningful outcome for the student, for example, in terms of an engaging problem. Utility is the construction of meaning for the ways in which mathematical concepts are useful. Ainley et al. (2006, p. 25) consider “the provision of authentic tasks inherently problematic,” but we think that suitable authentic contexts can provide both purpose and utility. The approach we investigate in this chapter is to base tasks in upper secondary education on problems from authentic professional practices. Research has shown that such a design approach can help students see the purpose of what they do in classrooms and the utility of what they learn. However, such an approach might come at the expense of conceptual learning.

A. Dierdorp • A. Bakker (✉) • J. van Maanen • H. Eijkelhof
Freudenthal Institute, Utrecht University, Utrecht, The Netherlands
e-mail: a.bakker4@uu.nl

The aim of this chapter is to examine whether realistic tasks, inspired by authentic professional practices, and embedded in appropriate teaching, can in principle support students’ conceptual understanding of sampling in such a way that they also learn to shuttle between contextual and statistical spheres. To us, this “shuttling” not only includes students’ application of statistical knowledge, but also seeing the purpose of the task, and the utility of what they learn.

4.2 Concepts Underlying Sampling

Basing tasks on situations from authentic professional practices has been studied in science education, but much less so in statistics education (Dierdorp, Bakker, Eijkelhof, & van Maanen, 2011). Here we focus on sampling to help students see the utility of correlation and regression and shuttle between the contextual and statistical spheres. Sampling is considered a key aspect to the teaching of informal inferential reasoning. For example, Pfannkuch (2008, p. 1) argued: “when students are not aware of sampling their informal inferential reasoning is limited.” We address five concepts underlying sampling (inspired by Pfannkuch, 2008, p. 4) which are important for students’ statistical reasoning: sample size, random process, distribution, intuitive confidence interval, and relationship between sample and population. We address these five concepts one by one and show later that we recognize these concepts in students’ reasoning. The chapter’s focus is schematically represented in Fig. 4.1.

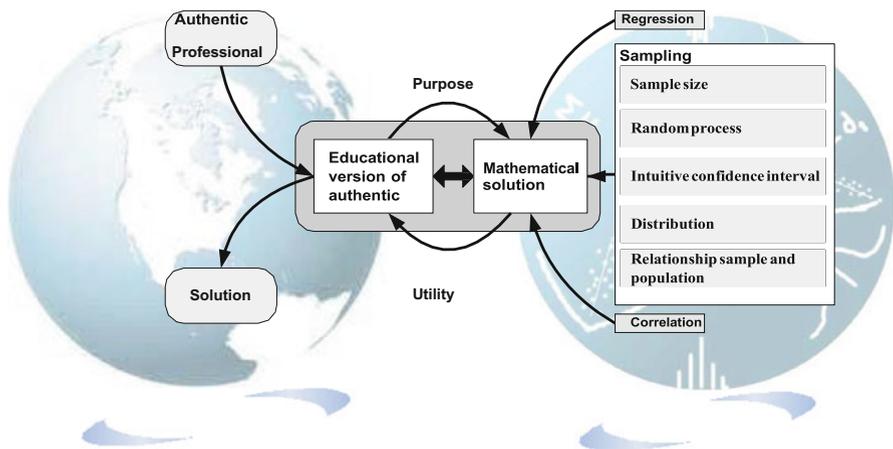


Fig. 4.1 Shuttling between the contextual and statistical spheres during the sampling tasks

4.2.1 Sample Size and Law of Large Numbers

Students need to understand that increasing sample size generally leads to better estimates of probability and population characteristics. For understanding sampling it is therefore necessary that students develop a concept of sample size. A “big idea” connected with sample size is the law of large numbers which says that predictions become more reliable when made from larger than from smaller samples.

4.2.2 Random Process

Students also need to realize that random processes such as repeated measurements of the same phenomenon will lead to different outcomes. Then they can understand that inferences are influenced by the drawn sample.

4.2.3 Distribution

The aforementioned concepts underlying sampling relate to the big idea of distribution. Rubin, Bruce, and Tenney (1990) argued that a sample gives the practitioner information about the distribution of a population and that this is a central idea of statistical inference. Drawing enough samples can support students’ understanding that the shape of the graph obtained by a “bigger” sample becomes more similar to the graph of the population as a whole. Confronting students with samples of increasing sizes, so-called growing samples, can help students to become more aware of emerging distributions by means of stabilizing measures of variation (tendency) and smoothening shape (e.g., Bakker, 2004).

4.2.4 Intuitive Confidence Interval

Shaughnessy (2006, p. 87) argues that students “should have a sense of the reasonably expected variability around the expected value, something as a confidence interval.” In real life, predictions are not based on one value obtained by a regression line because often a margin around the predicted value is essential. For example when a physiotherapist finds a client’s peak heart rate just under the value predicted by a common formula he or she will not worry (Dierdorj et al., 2011). He has a sense of what could be called an intuitive confidence interval.

4.2.5 Relationship of Sample and Population

Research into school statistics reports the problems students have in drawing inferences that make sense in the context (Makar & Rubin, 2009). One problem of how to draw sensible inferences is caused by a lack of awareness of variability when generating samples. Samples often provide a distorted image that is yet in some way representative.

4.2.6 Research Question

Given the problems and challenges mentioned in Sect. 4.1, and the complex multi-faceted concepts underlying sampling summarized in Sect. 4.2, we formulated the following research question: How can students be supported to develop concepts underlying sampling and shuttle between contextual and statistical spheres?

4.3 Method

The research question is addressed through two case studies that were part of a design experiment on correlation and regression with a twelfth-grade group of 13 students and an eleventh-grade group of 16 students from the preuniversity track. Both groups had opted to study the school subject “Nature, Life, and Technology.” The first group was taught by the first author. The second group was taught by another teacher, with the first author observing and interviewing. Two sample tasks covered three of the 23 lessons, each 50 min in both schools. Students used Fathom for drawing samples and Excel for making their own scatter plots to investigate their results from the sampling software.

4.3.1 Sampling Tasks

To stimulate students to develop concepts underlying sampling and to shuttle between contextual and statistical spheres, we designed two realistic tasks based on authentic professional practices.

To stimulate students’ reasoning about sampling we drew on a professional practice, namely, research on peak heart rates (PHR). Gellish et al. (2007) measured many people and found a different relationship ($\text{PHR} = 207 - 0.7A$, with A as age) between age and peak heart rate than the one typically used in sport physiotherapy ($\text{PHR} = 220 - A$). We provided the students with their data set of 908 measurements. In the Heart rate task, we asked the students if we could do with a smaller sample: What smaller sample size would be sufficient to find a reliable formula that is close to the original formula based on Gellish et al.’s data set?

The task is based on the instructional idea of growing samples. Such an approach is helpful in supporting coherent reasoning about sampling, distribution, and other statistical key concepts (Bakker, 2004). The students started with Gellish et al.'s data set and regression line, and then investigated smaller samples to find the smallest sample size that still produced a reliable model using the sampling option in Fathom. We conjectured that the question about how small a sample can be and still allow a sufficiently reliable inference can stimulate students to reason about all five aforementioned sampling concepts (and not just one aspect of sampling at the time, as happens in atomistic approaches to task design).

The second task is inspired by the practice of monitoring the height of dikes. A dike is an artificial construction to prevent flooding. Dike monitoring is essential for the Netherlands because large parts of the country are below sea level. A persistent problem is that dike heights decrease over time. If the height reaches a "critical value" (so named by the Ministry of Transport and Water Management) high sea and river water levels are a danger.

We introduced the Dike sampling task: Keeping in mind the high cost of measurement, what smaller sample would still have led to a reliable prediction of when the critical value will be reached? The students got 44 real data points of the deviations of a dike location.

The students could change the sample size and got the corresponding scatter plot with a regression line and a formula. They had to decide themselves which number of measurements was required to find a reliable prediction based on a smaller sample which is close to the prediction obtained by the regression line for the complete set. The students were only told that they had to save money by reducing the number of measurements.

To address the research question, we sought an efficient way to check if our tasks could in principle support students' concepts underlying sampling and their shuttling between context and statistics. To this end, it seemed sufficient to use case studies of students working with these tasks.

The first case study focused on Rick, aged 17. While Rick worked on the tasks, we video recorded his activities and transcribed the spoken text. To identify which sampling concepts were at stake in the interaction, we divided the transcripts into 26 fragments of interactions between students or between student(s) and teacher. Each fragment consisted of several turns, that is, the spoken text of a person which is not interrupted by another person as a turn. Each fragment was coded by means of interpretive microanalysis with one or more items of the five aforementioned concepts underlying sampling.

The second case study focused on two 11th-grade students, Sean and Kars, both aged 17 from another yet similar school. Sean and Kars were both motivated students, and their school results were similar to Rick's. Sean and Kars wrote in their questionnaires that they were not familiar with the authentic professional practices or the statistical techniques central in the tasks. This second case study focuses on the Dike sampling task, during which the researcher observed and interviewed the students. We video recorded the interaction and transcribed the spoken text. To analyze what teachers may need to do to stimulate students to shuttle

between the contextual and statistical spheres, we divided the transcripts into three phases and categorized the researcher's attempts to help them shuttle back from statistics to context.

4.4 Results

We present the results of the analysis of the data collected during the Heart rate and Dike sampling tasks and show what kind of reasoning about sampling these tasks potentially support.

4.4.1 *Developing Concepts Underlying Sampling (Case 1)*

Case study 1 was carried out to investigate the potential of the tasks to support students' development of concepts underlying sampling. We first present a quantitative impression.

From the rather well-balanced distribution of codes across the different concepts underlying sampling, we conclude that the tasks have the potential to raise students' concepts underlying samples. Many fragments were coded by more than one concept suggesting that these concepts were developed and used in relation to each other. In the following subsections, we illustrate the five conceptual concepts underlying sampling. The origin of excerpts is indicated by Fn, where n is the number of the fragment.

4.4.2 *Sample Size*

From the eleven fragments in sample size column of Table 4.1, we infer Rick had developed a sense of sample size and law of large numbers. For example, he expected that when the sample size becomes big enough the regression coefficients would stabilize. He also formulated it reversely as follows:

Rick [F4]: You get [when drawing 100 instead of 500 cases] more variability. More deviations but the line will have a negative slope. The intersection with the vertical axis will be different too. By changing the number of cases you can monitor the effect on the regression line.

This last sentence may indicate that he understood the utility of sampling. In other words, he seemed to understand why the concept of sampling is useful in this task. When he drew samples of size 50 he mentioned the difference of the coefficients of the regression lines when executing samples of this same size. He mentioned that when he drew bigger samples, he would expect a formula closer to the original formula of the population. Such examples indicate that Rick was aware of the effect of sample size and had a sense of the law of large numbers.

Table 4.1 Concepts underlying sampling recognized in each fragment of Rick's spoken text

	Fragment	Sample size	Random process	Distribution	Confidence interval	Relation sample/population	Time (s)	# Turns
Heart rate task	1	x					50	6
	2	x	x				36	5
	3			x			41	4
	4	x					24	3
	5	x		x			48	5
	6					x	56	9
	7	x	x				80	8
	8			x		x	56	7
	9	x		x			48	5
	10				x		71	9
	11		x				59	15
Disk task	12					x	13	2
	13	x	x	x			46	6
	14	x			x		67	6
	15	x			x		65	6
	16					x	20	4
	17		x	x		x	55	7
	18				x		15	3
	19			x		x	35	4
	20				x		20	4
	21					x	15	4
	22		x			x	15	2
	23		x	x			25	5
	24	x				x	40	7
	25	x		x	x		40	7
	26					x	50	6
	Total	11	7	8	7	12		

4.4.3 *Random Process*

There were indications that Rick developed understanding of the concept of random process during his work on the task. At first, he believed that a bigger sample size would lead to a more reliable regression line, but when he took a sample of size 100 he noticed that the regression coefficients deviated more than those of the last sample of size 50. He then realized that a relatively small extension of sample size would not necessarily lead to a “better” formula and tried to explain this with the concept that the software randomly sampled the higher PHR. The teacher tried to stimulate a reflection on Rick’s former statement that a higher amount of cases would imply a formula for the regression line more similar to the original. Rick seemed to be clear about the fact that the randomness of the sample leads to unpredictable outcomes and that it is not necessary [F7] to find a regression line more similar to that of the larger set when taking a slightly bigger sample. We see this as growth in Rick’s understanding of how samples behave.

4.4.4 *Intuitive Confidence Interval*

Rick mentioned that with bigger samples the points were getting closer together. He entered all regression coefficients in a spreadsheet, seemed to be aware of an interval per sample size, called this “margin,” and found a margin for samples size of 200 and up acceptable to make a prediction. Rick considered a margin around an expected value: “Till 200 it is too varied, if you do not combine. It can be a coincidence. You can build a safety margin. Then you go to 250.” He explained that the margin of the slopes he found at size 200 could be small as the result of coincidence and thought that a sample size of 200 would allow him to predict a reliable regression line. When he said “if you do not combine” he probably meant that this is only the case when focusing only on the slope. For safety reasons he suggested a sample size which was one “step” bigger, such as 250. The fact that Rick mentioned a safety margin supports that the realistic context of this task may have provoked him to reason about the outcomes.

4.4.5 *Distribution*

In previous lessons the students discussed the role of a physiotherapist using a regression line to advice his clients. Rick and his fellow students decided that a client’s peak heart rate must be in a margin around the regression line. When the teacher asked Rick if he expected a larger correlation when drawing a larger sample, he answered negatively and explained:

Rick [F13]: When you start taking random values and do so the tenth time, taking random values, they are still random. Only the margin will become more colored

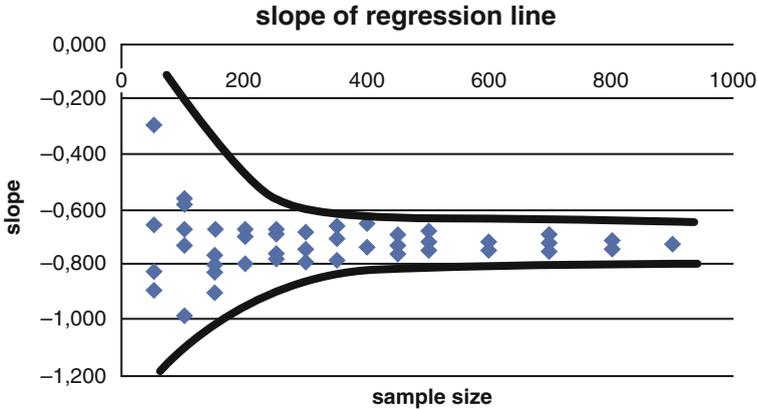


Fig. 4.2 The trumpet shape distribution of the slopes of regression lines obtained by growing samples drawn by Rick

[filled by colored dots]. It will not be wider and will never become narrower and never become much wider.

This excerpt indicates Rick’s sense of margin and distribution and that he seemed to be aware of the purpose of the task. He expected the margin to be about the same for each sample size. Only when drawing a sample with a bigger size did he mention that the margin became more colored. Probably, he mentioned this because he expected more points within the margin when the sample is larger.

Rick considered the shape of the distribution (see Fig. 4.2) which he called “a trumpet shape.” From his remarks it was clear that he had expected to see a stabilizing trumpet shape. He saw this shape in the Dike sampling task as well. This time he did not draw the trumpet but only mentioned: “You see again the trumpet shape.” At this point we proceed with the Dike sampling task, in which students were challenged to apply what they had learned in the Heart rate task.

4.4.6 Relationship Between Sample and Population

Rick showed some understanding of a relationship between sample and population and seemed to be aware of the purpose of the task: to find a smaller sample size in order to save money and still have a reliable formula for the regression line. He said to his fellow student Eline: “It [software] plots the regression line. Then you are able to see how much points [measurements] can be saved and still find a reliable regression line [F18].” Further analysis suggests that Rick was aware of the purpose of the task and the utility of sampling. Rick talking about the reliability of a sample with a certain size indicates that he thought about the relationship between the sample and population distribution. He compared the results of every sample with the results of the original set. The observations gave us indications that other

students also developed conceptual understanding of the relation between the sample distribution and population distribution, but shuttling back to the context was not obvious to them. Many only considered the trumpet shape distribution of the slope to consider the sample size for making a reliable inference. Only Rick and Eline shuttled between the spheres basing their inference on the total effect on the deformations instead of only looking at the slope of the regression line.

4.4.7 Using the Dike Sampling Task's Potential to Improve Students' Shuttling (Case 2)

To study how a teacher can use the task potential to improve students' shuttling between the statistical and contextual spheres, we carried out another case study. From this case study we present the Dike sampling task, in which the researcher needed to put a great deal of effort in supporting Sean to link back to the context. We distinguished three phases in Sean's working on the task with his fellow student Kars: Phase 1: in which students were self-reliant working on the task, Phase 2: with several types of questions and hints by the researcher, and Phase 3: in which the students were self-reliant working on the researcher's context question.

During the first 20 min (Phase 1), Sean and Kars worked self-reliantly. Although all tasks were based on authentic professional contexts and the students seemed to see the purpose of the task, they stayed in the math world focusing on formulas:

Sean: The slope $[-0.00124, \text{slope of the regression line at sample size } 20]$ is almost the same [as from the original formula, -0.00123].

Kars: Yes, the b too [the b from $y = ax + b$; -2.74 vs. -2.8].

After about 20 min, when the researcher discovered that Sean and Kars reasoned without referencing back to the contextual problem, he tried to focus their attention to the contextual meaning of their decisions in several ways (Phase 2). To gain insight into the support that the researcher gave, we classified his turns as informative questions, with which he tried to find out what the students did and meant ($n = 10$), revoicing questions and remarks in which he rephrased the students remarks ($n = 7$), explanation questions to find out why they did certain actions ($n = 5$), instructional support to do general suggestions ($n = 2$), and reducing degrees of freedom support to let the students shuttle between the contextual and the statistical sphere ($n = 1$). Despite the researcher's questions, Sean and Kars stayed focused on the slope of the regression line and did not use the context. Often the slope was different only in the fourth decimal, and they thought that the formula obtained by sample size 30 was close enough to the original formula. They judged this purely on the basis of the formula, whereas the researcher hoped they would think through the contextual consequences of the differences between the regression formulas.

At the end of phase 2, the researcher again tried to reduce the degrees of freedom and asked more specifically about the difference in days between the prediction based on the regression line with sample size 30 and the prediction based on the

original formula based on a sample of 44 (Phase 3). Sean and Kars then continued to work self-reliantly on this question. They used the spreadsheet to calculate the day for which according to the regression line $H=-10$, and found that the random sample with size 30 produced a regression line which would imply raising the dike almost a year later than the original line. Then they realized that the context asked for a more precise approach. Because Sean and Kars were no longer satisfied with sample size 30, they decided to take samples of other sizes. For each sample they also calculated the difference between the corresponding calculated predicted days and the day when the original formula would predict the critical value of -10 mm. They also calculated the average of differences for each sample size. They decided that in this context sample size 40 was acceptable. They were disappointed that they only saved four measurements, but when the researcher asked how much money would be involved in skipping four helicopter flights, they were more satisfied. This last case study indicates that the students were focused on the mathematical concepts of the tasks. The researcher repeatedly had to emphasize the contextual problem to stimulate the shuttling back to the contextual sphere. He had to reduce the degrees of freedom by asking about specific contextual consequences of difference in the formula obtained by their sample and the original formula.

4.5 Conclusion and Discussion

Our research question was how students can be supported to develop concepts underlying sampling and to shuttle between contextual and statistical spheres. It seems possible to use authentic problems from professional practices to design tasks that are purposeful from a student perspective and lead students to see the utility of what they learn. This might help students apply what they have learned. However, it is not self-evident that students develop rich conceptual understanding from authentic tasks because designers seem to have less control about what students learn conceptually. We wanted to know whether it is feasible, in principle, that students see purpose and utility while also developing a rich conceptual understanding of what is at stake, in this case sampling. In response we conclude that the analyses show that the realistic sampling tasks, inspired by authentic professional practices, are rich and focused enough to stimulate reasoning about the concepts underlying sampling in a balanced way and in relation to each other. This seems an advantage over atomistic approaches to statistics education that deal with aspects of concepts one by one (cf. Bakker & Derry, 2011) and this seems to address Ainley et al.'s (2006) concern that engaging tasks seem to be less focused.

Because with task design the devil is in the detail, it is worth speculating on what makes the sampling tasks suitable in helping students develop the main concepts underlying sampling in relation to each other. We think this has to do with the question of asking how small a sample can be so that the inference is still reliable. This explicitly requires reasoning about sample size and the relation between sample and population. Moreover, the issue of randomness comes up when students compare

samples with the same size when trying to judge if inferences are reliable. When comparing samples of the same size, intuitive confidence intervals and distributions also come into play. These findings are in line with earlier findings at the middle school level that tasks based on growing samples have the potential to stimulate students to reason about multiple facets of distribution and uncertainty.

We also wanted students to shuttle between contextual and statistical spheres. More concretely, we think it is important they learn to model contextual problems and apply their statistical knowledge, but also focus on purpose and utility. The purpose of the tasks was clear for the students involved: to find a smaller sample size in order to save money and still have a reliable formula for the regression line. It seemed that they did see the utility of sampling in order to find such a reliable formula.

By means of a second case study, we explored what types of questions teachers may need to ask. What helped best was to ask specific questions about practical consequences of mathematical issues (e.g., what would such a small difference in coefficients mean in terms of the prediction?).

We addressed a persistent design challenge and do not claim to have solved it. More research is needed to investigate the support students need to shuttle between contextual and statistical spheres. More specifically, we suggest investigating teachers' scaffolding of students' shuttling.

References

- 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). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64–83.
- Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, 13, 5–26.
- Dierdorp, A., Bakker, A., Eijkelhof, H. M. C., & van Maanen, J. A. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning*, 13, 132–151.
- Gellish, R. L., Goslin, B. R., Olson, R. E., McDonald, A. R. G., Moud, G., & Virinder, K. (2007). Longitudinal modeling of the relationship between age and maximal heart rate. *Medicine and Science in Sports and Exercise*, 39, 822–829.
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82–105.
- Pfannkuch, M. (2008, July 6–13). Building sampling concepts for statistical inference: A case study. In *Proceedings of the International Congress on Mathematical Education*, Monterrey, Mexico.
- Rubin, A., Bruce, B., & Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the 3rd International Conference on Teaching Statistics* (pp. 314–319). Voorburg, The Netherlands: International Statistics Institute.
- Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In G. F. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance* (pp. 87–98). Reston, VA: The National Council of Teachers of Mathematics.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67, 223–265.

Part II
Learning Statistics at the School Level

Chapter 5

The Language of Shape

Pip Arnold and Maxine Pfannkuch

5.1 Introduction

The power of statistical data analysis lies in describing and predicting aggregate features of data sets that cannot be noted from individual cases (Bakker, 2004, p. 100).

In a New Zealand national assessment, students in year 11 (age 15) are expected to be able to undertake a statistical investigation about a comparison situation. The assessment requires students to (in brief): pose an appropriate comparison investigative question; select and use appropriate display(s); give summary statistics; discuss features of distributions comparatively, such as shape; and communicate findings in a conclusion. For many years teachers have struggled with exactly what describing the shape of a distribution means and recent research on informal inferential reasoning (IIR) identified describing shapes of data distributions as an area where students demonstrated impoverished reasoning (Pfannkuch, Arnold, & Wild, 2011). Discussions were held with overseas experts and a clear solution was not evident, though fledgling ideas existed. These ideas were developed into activities to explore several aspects of distribution including the language of shape, making predictions and building a contextual knowledge base about shape.

P. Arnold (✉)
Cognition Education Ltd., Auckland, New Zealand
The University of Auckland, Auckland, New Zealand
e-mail: parnold@cognitioneducation.com

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

5.2 Literature Review

Over the last 10 years there have been a number of research projects with a focus on distribution and students' reasoning about distribution; for example, the Freudenthal Institute team (Bakker, 2004; Bakker & Gravemeijer, 2004), the Nashville team (McClain, 2005; McClain & Cobb, 2001) and the 2005 Fourth Statistical Reasoning, Thinking and Literacy Research Forum focused on reasoning about distribution. Five themes emerged from the research: (1) the notion of distribution, (2) measures of centre, (3) shape of distributions, (4) predicting distributions and (5) contextual knowledge. This chapter will focus on the shape of distributions and describing distributions.

Distribution is a multi-faceted notion involving centre, spread, skewness, shape and density (Bakker, 2004; Ben-Zvi & Amir, 2005; Konold, Higgins, Russell, & Khalil, 2004; McClain, 2005; Pfannkuch, 2005; Reading & Reid, 2006). Students need to consider measures of centre, measures of spread, where the majority of data values are in relation to extreme values, and how density and skewness provide detail about shape when viewing distributions. It is this global reasoning, the coordination of these ideas that makes distribution a complex notion that students find difficult (Ben-Zvi & Arcavi, 2001; delMas, Garfield, & Ooms, 2005; Hancock, Kaput, & Goldsmith, 1992; McClain & Cobb, 2001).

Describing shapes of distributions has had fleeting mention, with Bakker (2004) providing the only real depth in work on shape. Despite the relative superficial exploration of shape, there are some starting points to consider. Firstly, the type of graph used to display the data has a major influence on students' ability to perceive shape. For example, box plots and even histograms at earlier ages can prove a problem for students to use as they are too abstract and the actual data cannot be seen (Bakker, 2004; Friel, Curcio, & Bright, 2001). Dot plots, on the other hand, provide an initial starting point for students to explore shape along with simple case-value bar graphs (Bakker, 2004; delMas et al., 2005). Pfannkuch (2005) suggests that dot plots and stem-and-leaf plots can provide a strong basis for interpreting and understanding distributions and students can transition from them to box plots. Most critically displays used should allow sense to be made of the information with as much ease as is possible (Friel et al., 2001). Secondly, Bakker (2004) suggests that single univariate distributions are a good starting point, but cautions that students can initially assume that all distributions are symmetric if only this type is selected. Students' thinking can be challenged by deliberately choosing distributions that are skewed as well as symmetric (Bakker, 2004; delMas et al., 2005; Makar & Confrey, 2005; Rubin, Hammerman, Puttick, & Campbell, 2005). Linked to this is providing many opportunities for students to recognise and understand the direction of a skew (delMas et al., 2005), which is also a problem for college level students. Descriptors of shape include uniform, normal, skewed to the right or left (Bakker, 2004) and normal, skewed, bimodal or uniform (delMas et al., 2005), with early student ideas describing the data in terms of low, average and high values and naming shapes

using pyramid, semi-circle and bell shaped (Bakker, 2004). Thirdly, shape helps to develop meaning for mean, spread, density and skewness (Bakker, 2004; Rubin et al., 2005) and connections between measures of centre and shape can be made (Konold & Higgins, 2003; Rubin et al., 2005). Finally, Bakker (2004) found that too small a sample size, unsuitable scaling and lack of context were problems when trying to identify the shape of distributions.

An end goal is that students are able to describe sample distributions as part of the statistical enquiry cycle to answer an investigative question about a population. The research questions for this chapter are: What shapes do year 10 students (age 14) realise from data distributions? and what descriptions of distributions are year 10 students capable of producing?

5.3 Theoretical Frameworks

Two theoretical frameworks were considered in the analysis of student responses in pre- and post-tests. Bakker and Gravemeijer (2004) proposed a structure (Fig. 5.1) for analysing the relationship between data and distribution. They said that students as novices typically see individual values and use these to find values such as the median, range or quartiles, but that this does not mean they are seeing the median, for example, as representative of a group.

Ben-Zvi, Gil, and Apel (2007) IIR theoretical framework provides cognitive aspects that relate to distribution—reasoning about variability (spread, density), distributional reasoning (aggregate views, pattern and trend, hypothesis and prediction, individual cases, outliers), reasoning about signal and noise (centre, measures, modal clumps, summary), contextual reasoning (interpretation, alternative explanations) and graph comprehension (decoding visual shapes).

5.4 Methodology

The research method follows design research principles (Roth, 2005) for a teaching experiment in a classroom. In the preparation and design stage, the first author developed the teaching and learning materials to use in the teaching experiment in

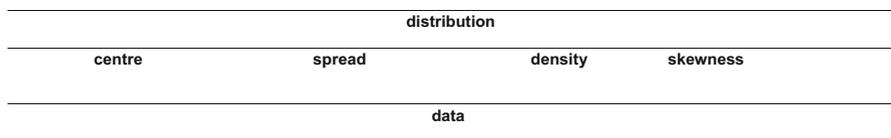


Fig. 5.1 Between data and distribution (Bakker & Gravemeijer, 2004, p. 148)

conjunction with the classroom teacher, considering relevant literature. In addition, purposefully built into the teaching and learning sequence were activities with a focus on shape prediction and building a ‘library’ of knowledge around contexts and shape but these are not reported in this chapter. Both the classroom teacher and first author were involved in the implementation of the activities in the teacher’s year 10 class. Following each lesson there was reflective discussion and adjustments were made as needed to the learning trajectory.

The 29 students in the class were above average in ability and from a mid-size (1300), multicultural, mid socio-economic inner city girls’ secondary school. Students were given a pre- and post-test, the lessons were videotaped and student work was photocopied. A group of six girls were observed specifically as well as the teacher led whole class discussions. The six girls also had pre- and post-interviews about their responses to their tests.

The retrospective analysis for this chapter focuses on the development of students’ use of the language of shape and their descriptions of distributions. The learning activities were designed to support students’ understanding of these two aspects. The activities built on work previously undertaken in IIR project (Pfannkuch et al., 2011). They also included new thinking as we considered the bigger picture of what we were trying to achieve. The new/updated activities were based on the themes that emerged from the literature: in particular, they focused on the language of shape, making predictions and building a contextual knowledge base for the sorts of variables that have symmetric, skewed or uniform distributions. Unpacking students’ existing contextual knowledge and misunderstandings were key ingredients in predicting distributions.

This chapter focuses on lessons 2–4 of a 16-lesson unit on statistics. Lesson one was a review. Lesson two had a focus on seeing and describing shape and involved developing the language for shape of distribution descriptors, sketching shapes from graphs, grouping similar shaped graphs and matching shape descriptors to groups of graphs. Shape of distributions was a big idea in the lesson. Lesson three had a focus on linking shape and context and involved making predictions of the graph from contexts, matching contexts to graphs and starting to develop a ‘library’ of similar shaped graphs. Big ideas in this lesson were shape of distributions, predicting distributions and contextual knowledge. Lesson four focused on using the language of shape to describe distributions and involved sorting graphs according to shape of distribution and starting to describe distributions. The big ideas in this lesson were the notion of distribution, shape of distributions and contextual knowledge. The first author (FA) taught lessons three and four as the teacher was ill.

5.5 Teaching Activities

The three lessons described demonstrate how student-generated concepts, ideas and language were gradually transformed towards a statistical approach.

5.5.1 Lesson 2: Seeing and Describing Shape

The students firstly sketched the shape of 15 data distributions that were briefly shown using a PowerPoint presentation. Secondly, the students grouped the sketches of the graphs into similar shapes and used their own language to describe the shapes in each group. At this point the teacher asked about the number of groups they had made. For example:

Teacher: Four groups, what were they based on?

Student: Sloped to the left, and sloped to the right, symmetric ones

Teacher: So you have sloped to the left, sloped to the right, symmetric and what was your other group?

Student: You know [gestures with hand—up, across and down] it is even on the top

Teacher: Even on the top, so let's see, symmetrical, some sloped to the left, slope to the right, other one was ... [Various student responses with 'flat top' being the loudest]

The teacher used these four group headings—symmetrical, sloped to the left, sloped to the right and flat top—as a starting point. The class then sorted the graphs into one of the four groups (Fig. 5.2). Finally, the students were introduced to the statistical language used to describe shapes and were asked to match these words to their graphs. Intuitively the students re-grouped the graphs according to symmetry, symmetric or not symmetric, splitting the skewed into two groups (left and right) and the symmetric into two groups (uniform and other). Interestingly modality was not used for grouping.

5.5.2 Lesson 3: Linking Shape and Context

In order to get students to think about how context and shape were linked, they were given 15 contexts without the graphs and asked to sketch a shape for these contexts with some possible values. Discussion justifying shapes for particular

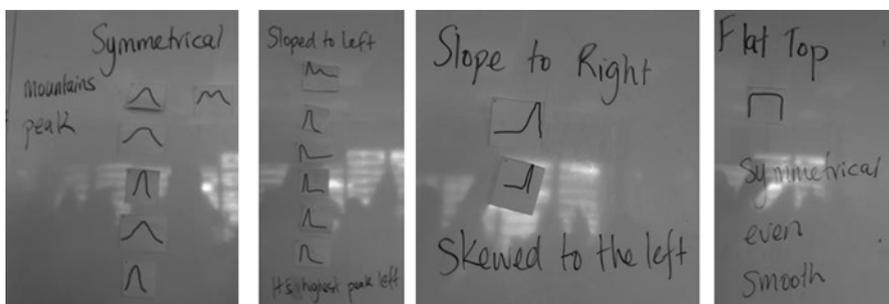


Fig. 5.2 Four shape groups with graphs and additional ideas

contexts followed. Students were then given the actual dot plots of the contexts and they matched these plots to the context. The graphs were sorted again into the four groups (symmetric, sloped to the left, sloped to the right, flat top) and each group was labelled using appropriate statistical terminology—symmetrical, right skew, left skew and uniform, including a discussion around why and which way the skew was recognised. At this point the distinction between unimodal and bimodal was made.

FA teacher: These are the graphs yesterday that you said were symmetric, and I've moved this one out to the bottom. Why do you think I have done this?

Student: Because it's bimodal.

FA teacher: Because it's bimodal. So these are symmetric, and unimodal, which means that they have one bump or one peak. So they have one mode, or peak and this one here is symmetric and bimodal because it has two peaks.

From this brief conversation the way to sort the shapes became clear—sort by symmetry and then by modality (Fig. 5.3). The shape descriptors developed from the way the students intuitively sorted the graphs. The students did not make a separate group for bimodal, as the research team did. They sorted into four groups and then split three of the groups by modality.

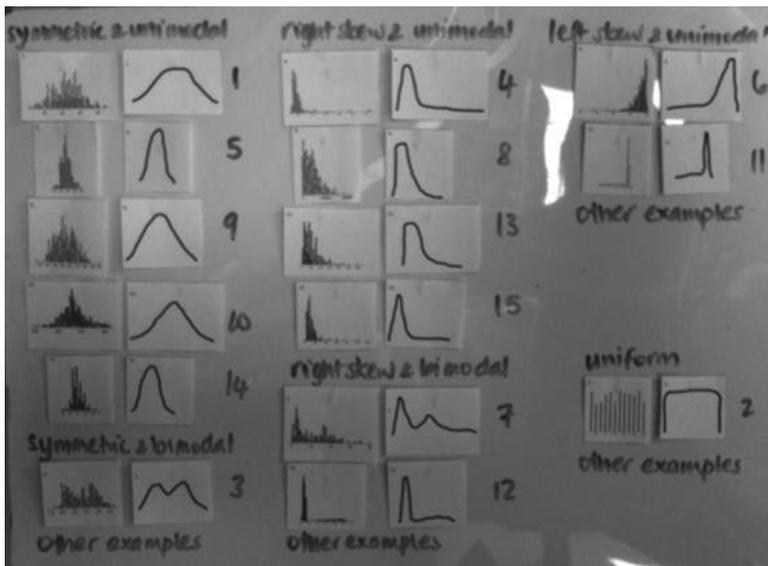


Fig. 5.3 Final collation of shapes into four groups with modality distinction

5.5.3 Lesson 4: Using the Language of Shape to Describe Distributions

The students firstly classified some more data graphs by shape and added these to their growing ‘library’ of shapes and contexts (note in Fig. 5.3 the label other examples). The next activity involved starting to describe distributions. The FA teacher facilitated a class discussion on key features for describing graphs. They were given the challenge that if they had to draw the graph from the description, what information would they need. Shape was a given, but a number of other features surfaced. A few excerpts from the discussion are:

FA teacher: So if I was going to describe this graph what other things might I want to describe about it?

Student: The range.

FA teacher: What other things would be important?

Student: Its highest point.

FA teacher: What are we calling that highest point?

Student: The peak....

FA teacher: What else might we want to talk about? What makes that graph (points to number 3) different to say number 14 (see Fig. 5.3)?

Student: The amount of peaks [modality].

In the discussion the features that the students suggested included: target population, variable, units, general shape sketched, overall shape, modality, peaks, range, median and mode. A further conversation in the same lesson where the focus was on describing one of the right skew graphs additional features surfaced: clustering density, majority, modal group and describing shape in terms of parts of the whole.

5.6 Pre- and Post-test Written Responses

Student pre- and post-test responses were analysed to see if their ability to describe distributions had improved over the course of the statistics unit. In one of the questions, students were asked for each of three situations (see Fig. 5.4) to *sketch the*

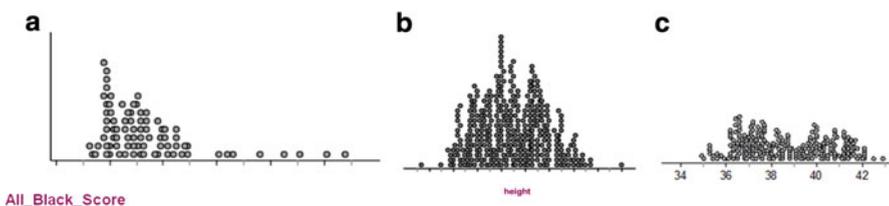


Fig. 5.4 (a) All Blacks’ (NZ rugby team) scores in test matches 2005–2010; (b) heights of NZ year 5–10 students; (c) heights of Tokoeka Kiwis (NZ native bird)

shape of the distribution of the variable and to write two statements about the distribution of the variable.

The SOLO taxonomy (Uniservices asTTle Team, 2008) was used as a basis for grading student responses. The particular descriptors aligned to each question were developed through a process of moving between the literature, in-class observations and student responses. In brief the descriptors for grading the student responses are: *no response* (NR-0); *pre-structural* (PS-1)—context and/or evidence missing; *uni-structural* (US-2)—gives one correct piece of evidence in simple context OR multi-structural evidence without any context; *multi-structural* (MS-3)—identifies a simple context and correctly describes two features OR relational evidence without any context; *relational* (R-4)—identifies the context, connects the context and correctly describes the overall shape and at least two other features; *extended abstract* (EA-5)—identifies the context, connects the context throughout the description, correctly describes the overall shape and at least three other features and may include some explanation or interpretation of results to the context (see Fig. 5.5c for an example of an extended abstract response).

The median grade across the three situations was used to represent the students' overall grade. These are summarised in Fig. 5.5a. In the pre-test the highest median grade was multi-structural with two students achieving this. In the post-test three students achieved at an extended abstract level and all but two students reached at least a multi-structural level. This means that the students could identify the context and describe at least two features. A lot of these students actually described more than two features, but they failed to make the broader link to the context, which was required to show relational thinking. The biggest movements were from students who scored 0–2 in the pre-test, perhaps indicating that acquisition of language and knowledge for describing distributions assists students. Figure 5.5b shows the median difference between students' pre- and post-test scores. The students made a significant improvement (p -value ≈ 0) in their median scores from the pre- to post-test question and on average increased their median grade by 1.7 points (95 % C.I. = [1.34, 2.07]).

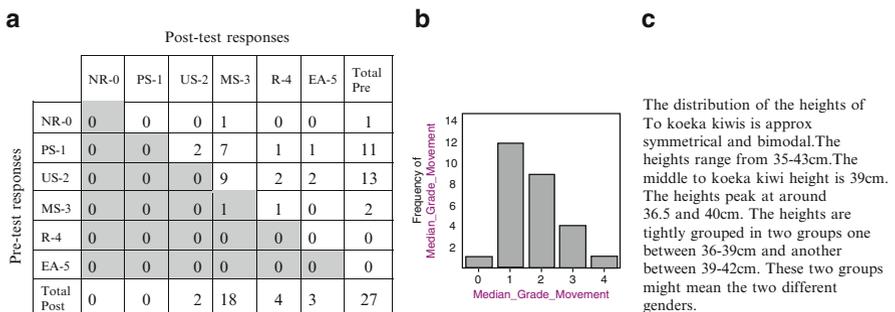


Fig. 5.5 Pre- and post-test results for one assessment question. (a) Pre- and post-test results. (b) Median grade movement from pre- to post-test. (c) Extended abstract response example for plot in Fig. 5.4c

5.7 Conclusions

‘Distribution’ is another fundamental given of statistical reasoning. I can find a great deal written about specialised usages and definitions of ‘distribution’ but almost nothing about ‘distribution’ itself as an underlying conceptual structure (Wild, 2006, p. 10).

The research questions were: What shapes do year 10 students (age 14) realise from data distributions? and What descriptions of distributions are year 10 students capable of producing? The year 10 students in this study intuitively sorted the data distributions into four groups, symmetric, right and left skew and uniform. These groups were further refined using a modality distinction. The classification realised by the students was based on what they noticed as they sought to group the graphs by shape. The teacher acknowledged student language and introduced appropriate statistical language which was connected to their four groups. These year 10 students appear to have the capacity to write thorough descriptions of data distributions. Further work and teacher modelling is needed to move students to a relational level where they can see the significance of parts of the whole description and intertwine context throughout the description.

Distribution is a complex notion. During the retrospective analysis phase, when student pre- and post-test responses were analysed, the two frameworks (Bakker & Gravemeijer, 2004; Ben-Zvi et al., 2007) that had previously been considered were found to only provide part of the picture. These frameworks need to be linked with a specific focus on the underlying conceptual structure of distribution. Combining the two frameworks led to our new proposed framework, the Distribution Description Framework (DDF), for thinking about, exploring and describing distribution. The DDF (Fig. 5.6) is organised by: (1) *overarching statistical concepts* that underpin distribution, (2) *characteristics of distribution* and (3) the *specific features* that are used when describing distributions.

Ben-Zvi et al. (2007) cognitive aspects from their IIR theoretical framework—reasoning about variability, distributional reasoning, reasoning about signal and noise, contextual reasoning and graph comprehension—were used to inform the overarching statistical concepts for distribution descriptions. Bakker and Gravemeijer’s (2004) characteristics of distribution—centre, spread, density and skewness—formed the backbone, with Pfannkuch, Regan, Wild, and Horton (2010) ideal data dialogue providing further characteristics and features to supplement those listed in the IIR theoretical framework. The result of the analysis of student pre- and post-test responses and in-class observations in this research provided the additional characteristics and features noted in Fig. 5.6 in italics.

Collectively these sources of data and ideas build a richer picture of the possible features that may be present in a particular distribution. While some aspects will be true and relevant in all descriptions (e.g. variable, overall shape), others (e.g. clustering density, mode) will depend on the data and whether or not they are relevant in the description. When students are describing statistical distributions they need: (1) to invoke contextual knowledge, (2) to know what relevant characteristics of

Overarching statistical concepts	Characteristics of distribution	Specific features measures/depictions/descriptors
Contextual knowledge	Population	Target population (e.g. NZ Yr 5-10 students) Other acceptable population (e.g. Yr 5-10 students)
	Variable	Variable Units
	Interpretation	Statistical feature described in contextual setting (e.g. interpreting right skew as very few high test scores, with most test scores between 20-50 points)
	Explanation	Possible reason for a feature (e.g. bimodal due to gender for kiwi data)
Distributional	Aggregate view	General shape sketched Hypothesis and prediction
	Skewness	Position of majority of the data
	Individual cases	Highest and lowest values
Graph Comprehension	Decoding visual shape	Overall shape *Parts of the whole (splitting the distribution into parts and describing the parts as well as the whole) Modality
	Unusual features	Gaps Outliers
Variability	Spread	Range, inter-quartile range *Interval for high and/or low values (may be describing a tail)
	Density	Clustering density Majority (mostly, many) Relative frequency
Signal and noise	Centre	Median, mean
	Modal clumps	Peak(s) (local mode) Modal group(s)

Fig. 5.6 Distribution description framework for year 10 (asterisk indicates part of feature listed)

distributions they can actually see in the plots and therefore describe and (3) to be explicit about the evidence for specific features. In other words, students need to be able to identify which features are evident in a particular plot, name and provide evidence (values) for the features and to interlace these with contextual information such as the population, variable and units.

Bakker and Gravemeijer’s (2004) framework appears to be about data distributions. In this study, the data distribution is conceived as a sample distribution and therefore more concepts come into play. At year 11 students are introduced to new concepts such as sampling variability, sketching inferred shapes and comparing sample distributions. This means that the DDF would be extended with students co-ordinating more ideas. The DDF would expand again in senior secondary where students start to consider distributions of statistics. Similarly, the DDF can be modified to support student progressions at lower curriculum levels. We believe our DDF has the potential to inform curriculum developers, researchers and teachers as they introduce students to the conceptual structure underlying distribution. Further research is needed both above and below the level reported here to ascertain what is appropriate for students at the different levels.

References

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: Freudenthal Institute.

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer.

- Ben-Zvi, D., & Amir, Y. (2005, July 2–7). *How do primary school students begin to reason about distributions?* Paper presented at the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4), Auckland, New Zealand.
- 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., Gil, E., & Apel, N. (2007, August 11–17). *What is hidden beyond the data? Helping young students to reason and argue about some wider universe.* Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5), University of Warwick, Coventry, England.
- delMas, R., Garfield, J., & Ooms, A. (2005, July 2–7). *Using assessment items to study students' difficulty reading and interpreting graphical representations of distributions.* Paper presented at the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL), Auckland, New Zealand.
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124–158.
- Hancock, C., Kaput, J., & Goldsmith, L. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 2(3), 337–364.
- Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.
- Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2004). *Data seen through different lenses.* Amherst, MA: Scientific Reasoning Research Institute, University of Massachusetts.
- Makar, K., & Confrey, J. (2005, July 2–7). *Using distributions as statistical evidence in well-structured and ill-structured problems.* Paper presented at the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL), Auckland, New Zealand.
- McClain, K. (2005, July 2–7). *The evolution of teachers' understandings of statistical data analysis: A focus on distribution.* Paper presented at the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL), Auckland, New Zealand.
- McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics*, 45, 103–129.
- Pfannkuch, M. (2005, July 2–7). *Informal inferential reasoning: A case study.* Paper presented at the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL), Auckland, New Zealand.
- Pfannkuch, M., Arnold, P., & Wild, C. (2011). Statistics: It's reasoning not calculating. Summary research report on Building students' inferential reasoning: Levels 5 and 6. Retrieved from www.tlri.org.nz.
- Pfannkuch, M., Regan, M., Wild, C., & Horton, N. (2010). Telling data stories: essential dialogues for comparative reasoning. *Journal of Statistics Education*, 18(1), 1–38. Retrieved from <http://www.amstat.org/publications/jse/v18n1/pfannkuch.pdf>.
- Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal*, 5(2), 46–68.
- Roth, W.-M. (2005). *Doing qualitative research: Praxis of method.* Rotterdam, The Netherlands: Sense.
- Rubin, A., Hammerman, J., Puttick, G., & Campbell, C. (2005, July 2–7). *Developing models of distributions using Tinkerplots.* Paper presented at the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL), Auckland, New Zealand.
- Uniservices asTTle Team. (2008). Quality questioning using the SOLO taxonomy. Retrieved from <http://www.slideshare.net/jocelynam/solo-taxonomy>.
- Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10–26.

Chapter 6

Development of an Understanding of a Sampling Distribution

Eun-Sung Ko

6.1 Introduction and Background

Sampling variability is generated because any two samples taken from a single population are not identical and because any single sample is not identical to the population (Franklin & Garfield, 2006). In sampling, the propriety of the sample size is not related to the size of the population. Sampling variability is affected by the sample size, not the proportion of the sample to the population. The greater the sample size, the less sampling variability there will be. Distribution is a tool through which we view patterns of variability (Wild, 2006). Reasoning about sampling variability requires an understanding of sampling distributions. The distribution of one very large-sized sample is similar in shape to its population. However, the shape of sampling distribution does not rely on population distribution. Studies have shown that students confuse the distribution of a sample and the sampling distribution and therefore they believe that the sampling distribution resembles the population distribution (Chance, delMas, & Garfield, 2004).

6.2 Method

6.2.1 Participants

The participants in this study consist of two groups of students: the first group is a group of mathematically talented students and the second group is a group of nontalented students. Table 6.1 shows a summary of the characteristics of the

This chapter uses data and analysis from the author's doctoral dissertation.

E.-S. Ko (✉)

Jeonju National University of Education, Jeonju, Jeollabuk-do, South Korea

e-mail: kes7402@jnue.kr

Table 6.1 Participants in this study

	Fifth grade	Eighth grade	Total
Talented students	34	36	70
Nontalented students	31	29	60
Total	65	65	130

participants of this study. At the time of this study the fifth-grade students had learned to display and summarize data, by working with tables, bar graphs, line graphs, stem-leaf graphs, and means. In addition, the eighth-grade students learned pie charts, frequency tables, frequency polygons, histograms, the relative frequency, the cumulative frequency, and probabilities.

6.2.2 Tasks

Figure 6.1 presents the tasks of this study. Q1 and Q2 are questions that were meant to examine students' thinking in relation to understanding of sampling distributions.

6.2.3 Data Collection and Analysis

Students were asked to complete questionnaires. The students for the interview were determined based on their responses on the questionnaire. They were students who presented typical responses in each category and students who did not fall into any category owing to their ambiguous languages and expressions. Interviews were video-taped or audio-taped, and transcripts of the interviews were made.

Two analyses were conducted. During the first analysis, students' responses to the questions were classified on the basis of their terms, languages, and expressions. These categories are not existing categories but are inductive products that emerged on the basis of their responses (Denzin & Lincoln, 1994; Goetz & LeCompte, 1984). During the second analysis, the structure of the observed learning outcomes (SOLO) taxonomy (Biggs & Collis, 1982) was used. Categories from the first step were regrouped into five groups considering the levels of the students' responses on the basis of the SOLO model. Both written responses and interview data were used to determine students' thinking levels. Intercoder reliability ($k=.874$) by two researchers was done.

6.3 Results and Discussion

Level 0. At this level, the students did not understand sampling variability. They believed that different samples have the same statistics. For example, they selected the first option, "Surveyor B obtains the same result as that of surveyor A" in Q1a

[Q1] The National Statistical Office examines the average size of the Korean family every two years. Surveyor A surveyed 1000 households and found that the average size of a family was four.

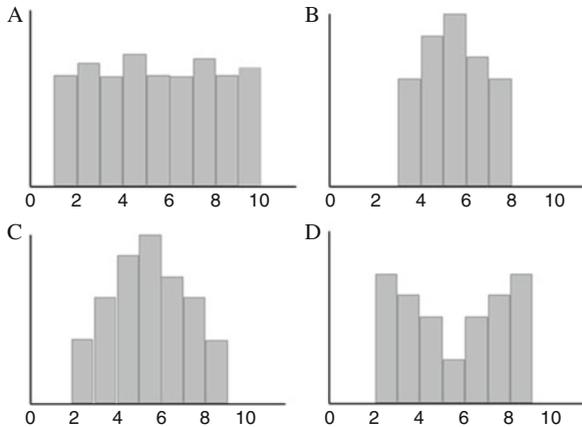
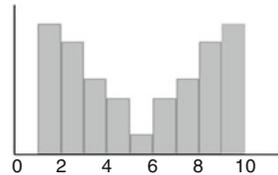
(a) Predict the result for surveyor B when he or she surveys 1000 households and select the best answer from the options below. Explain your selection.

- ① Surveyor B obtains the same result as that of surveyor A.
- ② Surveyor B obtains a different but similar result to that of surveyor A.
- ③ Surveyor B obtains an entirely different result from that of surveyor A.

(b) What do you think is the most appropriate method for determining the average size of a family? Select the best answer from the options below and explain your selection.

- ① The average size of a family is four because the survey produced four members per household.
- ② I will examine a set of data obtained from many 1000-household surveys.
- ③ I cannot determine the family size because results are different from survey to survey.

[Q2] The display on the right shows the numbers of pencils that each student in our country carries in their pencil case. (The horizontal axis indicates the number of pencils and the vertical axis denotes the number of students.)

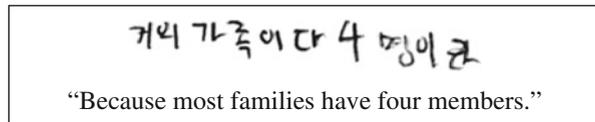


(a) Which display do you think shows the averages obtained from groups of four students? Select the best answer from A, B, C and D and explain your selection.

(b) Which display do you think shows the averages obtained from groups of twenty-five students? Select the best answer from A, B, C and D and explain your selection.

Fig. 6.1 Tasks of this study

Fig. 6.2 NE03's response to Q1b



and the first option, “The average size of a family is four because the survey produced four members per household” in Q1b. Below are some sample responses in level 0.

Figure 6.2 presents NE03's response to Q1b and the following transcripts are his explanations obtained during the interview. This student responded that there is no variability in the population based on his personal experience and that different samples have the same statistics.

Interviewer: You thought “B surveyor will get the same result as A surveyor” and “The average size of a family is four because the survey produces four members per household.” Why did you think so?

NE03: Because there are four members in my family and there are also four members in my friends' families.

Interviewer: Your friends and you have four-member families. If so, do all surveys obtain the same results?

NE03: Yes.

Level 1. At this level, the students confused the data in a sample with sample statistics as a data set. While they believed that different samples can have different statistics, they confused the data in a sample with sample statistics, and as such, did not recognize the patterns in the sample statistics. For example, they did not select the first option in both Q1a and Q1b but also did not select the correct option in either Q2a or Q2b. Below are some example responses in level 4.

Figures 6.3 and 6.4 present TE23's responses to Q2a and Q2b, respectively. This student selected the first graph in Q2a. He believed that sample statistics are different from population statistics because the sample size is too small. He also selected the last graph in Q2b. He believed the sample statistics has a similar distribution to that of the population because the sample size is big. He confused the data in a sample with the sample statistics. He focused on the likeness between the sample and the population but not on the patterns of the sample statistics.

Figure 6.5 presents NM12's responses to Q2a. This student selected the first graph in Q2a. He confused the sample statistics with the individual data. More specifically, he confused the number of pencils that each student carries with the set of the sample statistics. He also selected the last graph in Q2b. Figure 6.6 presents his responses to Q2b and the following transcripts are his explanations obtained during the interview. His responses show that he confused sample statistics with individual data. To understand sampling distribution, first of all, students should recognize the creation of a new data set composed of sample statistics and visualize the distribution of the new data. This student, however, did not understand the creation of a new data set composed of sample statistics.

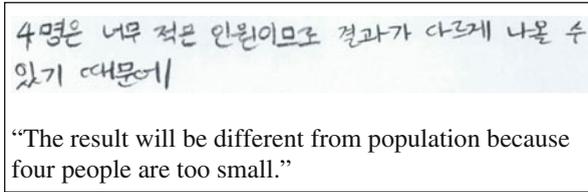


Fig. 6.3 TE23’s response to Q2a

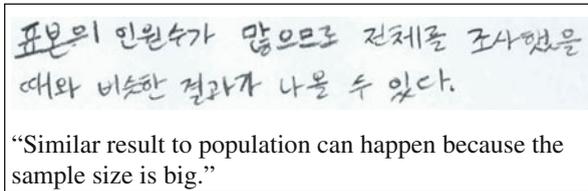


Fig. 6.4 TE23’s response to Q2b

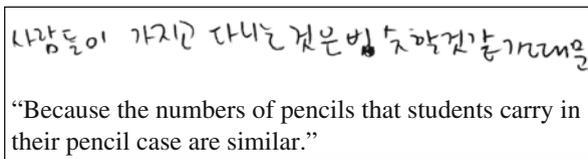


Fig. 6.5 NM12’s response to Q2a

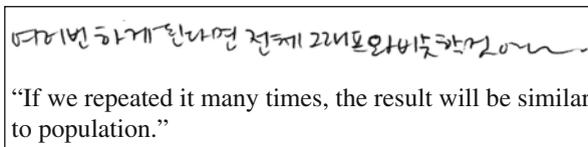


Fig. 6.6 NM12’s response to Q2b

Interviewer: Why do you think the result will be similar to the population if it was repeated many times?

NM12: If we do it many times, we can get as many data as the population.

Level 2. At this level, the students paid attention to the spread of the sample statistics but did not recognize their central tendency. Consequently, they did not infer that the center of sampling distribution parallels that of the population and did not visualize the correct shape of the sampling distribution. Below are some example responses in level 2.

TE31 selected the first graph in both Q2a and Q2b. Figures 6.7 and 6.8 show his responses to Q2a and Q2b, respectively, and the following transcripts are his explanations obtained during the interview. This student thought similar means are obtained when we calculate means of four-size samples and 25-size samples. In other words, he recognized the spread of the means but failed to predict how they will be distributed.

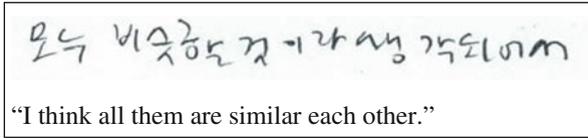


Fig. 6.7 TE31's response to Q2a

Fig. 6.8 TE31's response to Q2b

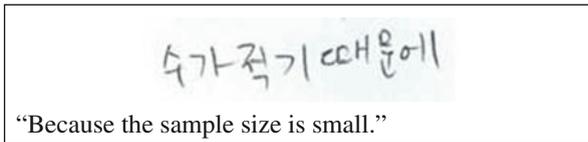
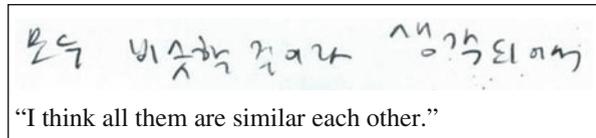


Fig. 6.9 TE16's response to Q2a

Interviewer: You wrote here “all of them are similar to each other.” What did you mean by “all of them”?

TE31: When we calculate the means of the four students, the means will be similar. And here, the means of the 25 students will be similar, too.

Level 3. At this level, the students recognized the central tendency and spread of the sample statistics. However, they did not understand the relationship between sample size and sampling variability. Below are some example responses in level 3.

TE16 selected the second graph in Q2a and the third graph in Q2b. Figure 6.9 presents TE16's responses to Q2a and the following transcripts are his explanations obtained during the interview. This student did not provide any written response to Q2b. He recognized the shape of the sampling distribution and the mean of the sampling distribution as six. However, he did not understand that the bigger the sample size, the lesser the sampling variability.

Interviewer: Here [in Q2a], you wrote “because the sample size is small” but here [in Q2b], you did not provide any response. Can you explain your answers for both questions?

TE31: First, let me explain [Q2b]. In finding out the means of the 25 students, each student carries different numbers of pencils. The means will have the most frequent six, and the more distant it is from six, the lesser the frequency.

Interviewer: How about [Q2a]?

TE31: It is similar [to Q2b]. If we calculated the means of four students, the means will have the most frequent six. And then the more distant it is from six, the lesser frequency.

Interviewer: The graph for the 25 students is different from the graph for the four students?

TE31: I think they are different because they have different numbers of students.

Level 4. At this level, the students recognized the central tendency and spread of the sample. They also understood the relationship between sample size and sampling variability. Below are some example responses in level 4. TM20 selected the third graph in Q2a and the second graph in Q2b. The following transcripts are his explanations obtained during the interview.

TM20: (Pointing to the population distribution) This shows a distribution of all the students. Now, we have to find out the means. I think, if we picked all the four-student samples from here (pointing to the bar indicating two pencils), the means will be around two. If we picked all four-student samples from here (pointing to the bar indicating eight pencils), the means will be around eight. However, six will have the most frequency. They will be gathering around six in the B and C graphs.

Interviewer: Why do you think six has the most frequency?

TE31: Because six is produced from four sixes, from two fours and two eights, from one four, one five, one seven, and one eight ... There are so many cases that produce six.

Interviewer: Why is the B graph for 25 students and the C graph for four students?

TE31: The 25 students are more [than the four students]. So it is more likely that the 25 students will be picked from the various groups. However, I think, for the four students, it is more likely to select the four students from one or two specific groups and produce extreme values. Considering the C graph, it has two and nine. For 25 students, the likelihood of two or nine is low.

Table 6.2 presents a summary of the characteristics of the students' understanding of sampling distribution at each level.

This study found that images of sampling and understanding of distribution are crucial for students to understand a sampling distribution. This study suggests that visual aids such as simulation can lead to instrumental understanding when students did not have images of sampling and understanding of distribution.

Table 6.2 Students' understanding of sampling distribution

Level	Students' understanding of sampling distribution at each level
0	No understanding of sampling variability: Students did not understand the concept of sampling variability. They believed different samples have the same statistics
1	Confusion of the data in a sample with the sample statistics: Students confused data in a sample with sample statistics, so they did not pay attention to the patterns of the sample statistics
2	Focusing on the spread of the sample statistics: Students recognized the spread but not the central tendency of the sample statistics
3	Focusing on the spread and center of the sample statistics: Students paid attention to both the central tendency and spread of the sample statistics
4	Understanding the relationship between sample size and sampling variability: Students recognized the central tendency and spread of the sample statistics and understood the relationship between sample size and sampling variability

References

- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The Solo taxonomy*. New York: Academic.
- Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295–323). Dordrecht, The Netherlands: Kluwer.
- Denzin, N. K., & Lincoln, Y. S. (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- 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. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance: Sixty-eight yearbook* (pp. 345–375). Reston, VA: National Council of Teachers of Mathematics.
- Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and qualitative design in educational research*. Orlando, FL: Academic.
- Wild, C. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10–26.

Chapter 7

Korean High School Students' Understanding of the Concept of Correlation

A-Ra No, Soo-Yun Han, and Yun Joo Yoo

7.1 Introduction and Background

The concept of correlation is one of basic statistical concept that can be used to understand the phenomena experienced in everyday life. The concept of correlation is related to covariational reasoning, reasoning about the relationship between two variables regarding their changes in values (Zieffler & Garfield, 2009). Correlation is a standardized measure of statistical covariation considering the variability of two variables observed in data. Covariational reasoning requires some understanding of bivariate data and the joint distribution of two random variables. The popular measure of correlation, Pearson correlation coefficient, is considered as an advanced topic in statistics since it requires the theory of joint distribution and the formula for Pearson correlation coefficient is difficult to understand intuitively.

Many countries such as United States, Australia, England, and New Zealand have been teaching the concept of correlation, time series, and linear regression along with related graphical representations as a part of mathematics education (Moritz, 2004). The secondary curriculum in Korea does not include the formula for Pearson correlation coefficient and only informal definition of correlation has been taught with bivariate data representations such as scatter plots and contingency tables.

Scatter plots and the concept of correlation are usually introduced together to students. Students learn about the data patterns of “correlation/no correlation” and “positive/negative correlation” for scatter plots. However, there could be some misconceptions generated since the concept of correlation is taught with only an informal definition. Cleveland, Diaconis, and McGill (1982) conducted an interesting study with two scatter plots of same data in that the scale of one plot is doubled,

A.-R. No • S.-Y. Han • Y.J. Yoo (✉)
Seoul National University, Seoul, South Korea
e-mail: callus01@snu.ac.kr; ssu1205@snu.ac.kr; yyoo@snu.ac.kr

which results in data points being more concentrated in the middle. The significant portion of subjects did say that the scatter plot with more concentrated data points looked more correlated even though they were told the correlation coefficients are same.

Contingency tables of two qualitative variables, especially 2-by-2 tables are introduced to teach the concept of correlation. Judging association from 2-by-2 tables require understanding of conditional probabilities (Baterno, Estepa, Godino, & Green, 1996). Without knowing formal test methods for the association in 2-by-2 tables, a person can judge the amount of association informally if she understands the comparison of conditional proportions at each level of one variable can give a brief idea on how two variables are related.

Also, the correlational relationship can be described in a text within scientific literatures, newspapers, or magazines, which requires correct interpretation based on the knowledge of correlation (Gal, 2004). Students may need the knowledge of correlation when they are learning other subjects of science and social studies.

As discussed above, the concept of correlation is inherent in various forms of data from everyday life. Because the ability to identify relationships between different factors through the concept of correlation is required in various fields (Zieffler & Garfield, 2009), many countries teach correlation in secondary education (Moritz, 2004). Since the 2007 Revised National Mathematics Curriculum,¹ however, Korean school mathematics does not cover the concept of correlation not even informally in the secondary education (Ministry of Education, Science and Technology, 2007). Therefore, a question can be raised: Can Korean high school students interpret various forms of statistical data and judge correlations among variables in the data without learning about correlation or with learning it informally?

To find answers to above question, this study set research questions as follows:

1. Are Korean high school students who have learned the concept of correlation informally able to understand correlation-related characteristics clearly in a scatter plot? What misconceptions arise from the visual characteristics of scatter plots?
2. Are Korean high school students who have learned the concept of correlation informally able to interpret correlations correctly in texts given in the context of real-life situations?
3. Are Korean high school students who have learned the concept of correlation informally able to interpret a 2-by-2 contingency table and tell whether there is a correlation between two qualitative variables and to present reasons for their judgment?

¹ 2007 Revised National Curriculum was announced on 28 February 2007, and this curriculum has been implemented since the 2009 school year starting with the first year students in middle school (seventh grade) and the first year students in high school (tenth grade). The subjects of this study, who are the first- and second-year high school (tenth and 11th grade) students at the time of 2011 had been taught under the previous curriculum, seventh curriculum in their middle school period. These students learned the concept of correlation informally through the bivariate data presentations of scatter plots in the middle school.

7.2 Method

7.2.1 *Research Subjects and Methods*

This study surveyed Korean high school students' understanding of the concept of correlation in order to find implications for Korean secondary mathematics education, in particular, statistics education. For this purpose, we developed items to examine the understanding of the concept of correlation. Items ask about data given in three different forms: scatter plot, 2-by-2 contingency table, and text stating correlational relationship of two variables.

In order to refine items and check the validity and the reliability of the test, we first asked seven experts' opinions on items in terms of content validity and then conducted two pilot studies. The first pilot study was performed on 47 tenth-grade students in a high school in Incheon. The second pilot study was performed on 40 tenth-grade students in a high school in Seoul. Based on the analysis of items with the data from pilot study, we finalized a test consisting of 25 items for 20 min administration time. The main study was conducted on 140 tenth-grade students and 100 11th-grade students from two high schools in Seoul. These students have learned about scatter plot, correlation, and contingency table in ninth grade. As insincere responses could distort the results of data analysis, they were excluded from the analysis, based on the number of missing values and extreme responses. The most frequent type of insincere respondents was those who did not reply a large number of questions. Respondents who did not answer eight (around 23 % of the whole) or more of the 25 items were all removed and, as a result, 19 students were removed. Then, five extreme respondents who chose the same alternative for all the questions were also removed. Resultantly, 24 students were removed out of a total of 240 respondents and analysis was made with data from 216 respondents.

7.2.2 *Categories and Contents of the Items*

As presented in Table 7.1, the test instrument consisted of three categories in terms of the tools of presentation for the correlation in the bivariate data: scatter plot (15 items), text (3 items), and 2-by-2 contingency table (7 items). Of the 25 questions in the questionnaire, 22 are of multiple choice type and three of true–false type. For ten items, students were asked to write the reason for their choices. The reliability of the test was computed with SPSS and Cronbach's alpha was .645.

7.3 Result

7.3.1 *Correlations Presented in Scatter Plots*

Items 1–3 and 5 ask about distinguishing several scatter plots between positive/negative/no correlations. The correct response rates for items 1–3 and 5 were 93.52 %, 93.52 %, 84.26 %, and 88.43 %, respectively, all exceeding 80 %.

Table 7.1 The concept of correlation measured by the test items and the corresponding item numbers

Representations	Objectives	Items
1. Scatter plot	(a) Determines “correlation/no correlation” and “positive/negative” correlations from the data in scatter plot	1–3, 5
	(b) Compares the strength of correlations in two scatter plots	4 , 11–15
	(c) Misconception regarding the steepness of the slope of regression line and the strength of correlation	6, 7
	(d) Knows that the change of the scale of variables does not change the strength of the correlation	8
	(e) Knows the influence of outliers on correlation	9, 10
2. Text	Correctly interprets the correlation given in real-life situation example	<u>16–18</u>
3. 2-by-2 contingency table	(a) Compares the strength of the association between two populations from two 2-by-2 contingency tables	19
	(b) Determines “correlation/no correlation” from data presented in 2-by-2 contingency table with proper justification	20–25

The multiple choice items that require the reason of the choice in **bold** and the true–false items in underline

Most students were able to interpret the presence and direction of correlation in given scatter plots.

The items 4, 11–15 ask students to compare the strength of correlations among multiple scatter plots. The correct response rates for these items were 79–89 % except for item 13 (47.69 %). Item 13 asks to determine higher correlation between one plot with strong negative correlation and one with modest positive correlation whereas other items require comparing the strength of correlation between plots with same direction of correlation (or with no correlation). For item 4, students were asked to write the reason of the choice. Students determined the strength of correlation simply based on the density or dispersion of points in the scatter plot and only 7.87 % related the strength of correlation to linearity. This result suggests that students might have some confusion between the strength of correlation and the direction of correlation.

Items 6, 7 are for the misconception regarding the steepness of the slope of regression line and the strength of correlation. Students were shown four scatter plots presented in Fig. 7.1. In item 6, students had to find correct statements. A total of 26.85 % students chose the statement “Among plot (b)–(d), the weakest correlation is the plot (b)” as a correct one, revealing the misconception that the less steeper regression line means weaker correlation. In item 7, students compared the correlation between plot (b) and (c), which actually have the same correlation coefficient value with different slope coefficient in the regression line. Only 17.34 % of the students determined that their correlations are similar (correct response), and 72.22 % students thought the plot (c), the steeper slope, has stronger correlation

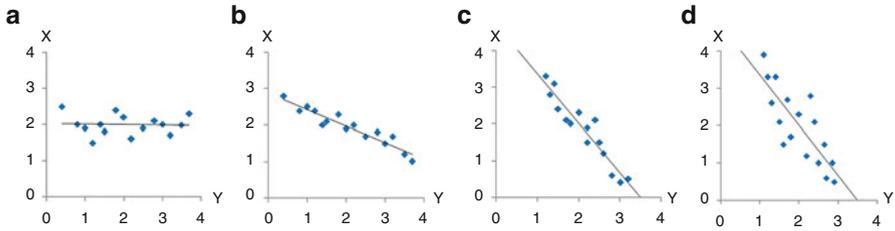


Fig. 7.1 Four scatter plots corresponding to items 6, 7

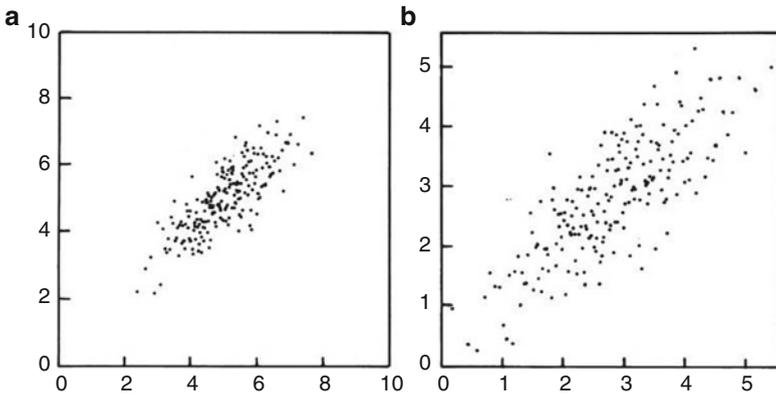


Fig. 7.2 Two scatter plots corresponding to item 8

than plot (b). For item 7, students were asked to write the reason for the choice. The students who have chosen plot as the stronger correlation stated the reasons such as “because the slope is steeper,” “because the change (of the regression line) is more obvious,” “because (c) is closer to the form of graph $y = -x$,” etc. All of these are directly or indirectly connected to the slope of regression line.

Item 8 was adapted from Cleveland et al.’s (1982) third experiment. Students were asked to compare two scatter plots of the same data with different scales (Fig. 7.2). In the explanations for the plots, it was clearly stated that the plot (b) is the enlarged version of plot (a). However, 60.04 % replied that scatter plot (a) has a stronger correlation, 18.52 % replied that scatter plot (b) has a stronger correlation, and 18.98 % replied that the two plots have the same correlation. These proportions were quite similar to the results of Cleveland et al. (1982), which were 66 %, 13 %, and 22 %, respectively. Students who had chosen plot (a) as the stronger correlation stated that “points in (a) are distributed more densely than those in (b).”

Items 9 and 10 were based on the questionnaire of Zieffler and Garfield (2009). A scatter plot was presented (Fig. 7.3) and students were asked about the change in the correlation if an outlier was removed at the location (a) (item 9) and if an outlier was added at the location (b) (item 10). The choices for the change of the correlation

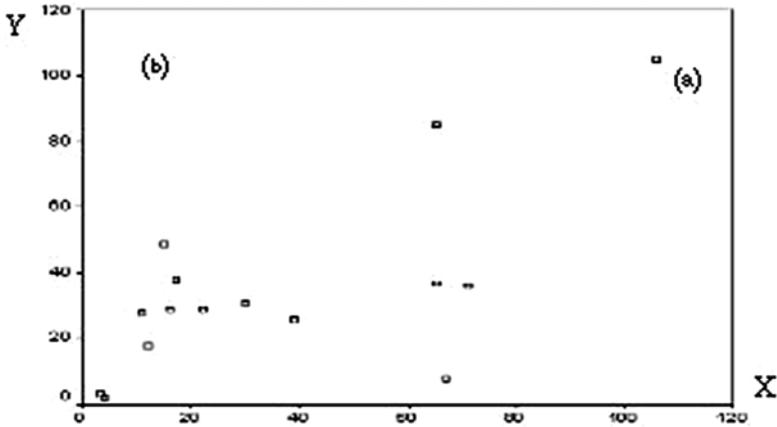


Fig. 7.3 A scatter plot corresponding to items 9, 10

were “becomes stronger,” “becomes weaker,” “no change.” The correct response rates of the items 9 and 10 were 46.3 % and 51.39 %, respectively. Also, 35.19 % and 36.57 % of the students, respectively, replied that the addition or removal of outliers did not change the correlation.

For true–false type items 16–18, a text describing the results of a statistical investigation was given as follows:

The Math and English exam scores of seventh-grade students and ninth-grade students in a middle school were investigated. We found that Math scores and English scores were correlated overall. Also, the correlation between Math scores and English scores were stronger in the ninth-grade students than the correlation in the seventh-grade students.

The statements for true–false type items 16–18 were as follows:

16. The relationship between Math scores and English scores is closer in the ninth-grade students than the correlation between Math scores and English scores in the seventh-grade students.
17. The differences in English scores for the students with Math score of 60–70 in the ninth grade were bigger than the differences in English scores for the students with Math score of 60–70 in the seventh grades.
18. The difference of English scores of two seventh-grade students whose Math scores are ten points apart is bigger than the difference of English scores of two ninth-grade students whose Math scores are ten points apart.

The correct response rates for items 16–18 were 88.43 %, 74.54 %, and 47.69 %, respectively. The intent of item 17 was to see if students understand less correlation means more dispersion in local distribution of one variable. In the responses to item 18, students also revealed confusion between the correlation and the slope of correlation. Some students tried to solve the problem by drawing scatter plots for the situation described by the text in scatter plots (Fig. 7.4).

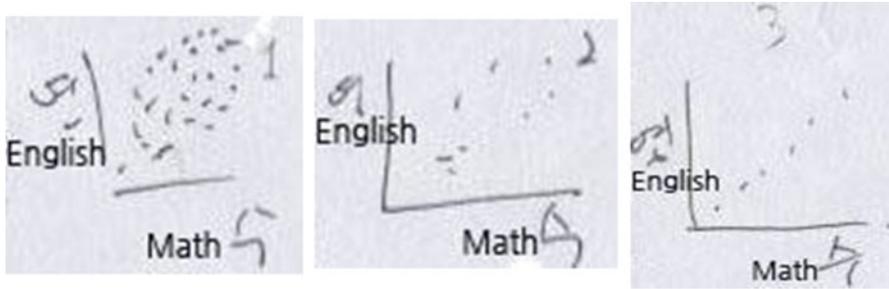


Fig. 7.4 An example of a drawing of scatters plot to understand the text

7.3.2 Correlations Inferred from 2-by-2 Contingency Tables

For item 19, two 2-by-2 contingency tables were shown and students had to compare the strength of association between two populations (Table 7.2). A total of 77.31 % of the students have chosen UK as showing stronger correlation between gender and the choice of cars. Students stated the reasons for the judgment such as “because inclination to a specific type of cars according to gender is stronger in the USA,” “because difference between men and women is more obvious in the UK.”

Items 20–25 were developed modifying the items of Baterno et al. (1996) in order to test whether the students can determine the presence of association between two qualitative variables presented in each of six 2-by-2 contingency tables. All the numbers of the 2-by-2 contingency tables presented in the items were verified for statistical association through Fisher’s exact test. Students had to state the reason for their judgment.

Associations were present in Table 7.3 (item 20) and Table 7.4 (item 21). The difference between two items was that the predictor was the row variable in item 20 whereas the predictor was the column variable in item 21. The correct response rates of two items were 75.00 % and 68.52 %, respectively. As the reasons for the answer to item 20, 156 of 162 students who answered item 20 correctly replied “because there is obvious difference in the presence of skin allergy according to lifestyle” and three of them replied “because those having sedentary lifestyle are exposed less to virus causing skin allergy.” These answers suggest that they replied not based on given data but based on their preconceived ideas. Such a pattern of response was observed for the reasons for the judgment to item 21.

The data in Table 7.5 (item 22) and Table 7.6 (item 24) had no correlation, and only 35.19 % and 57.41 % of students answered such, respectively. The difference in the correct response rate between the two items was statistically significant (p -value < 0.001). The difference of two tables was in that the sample sizes of predictor categories (row variable) were same in the item 24 but different in the item 22. Students with incorrect judgment might have compared the absolute numbers of the data instead of conditional proportions in each category of the predictor variable. In the reasons stated for item 22 and item 24, only 14.35 % (31, 35.19 % of the

Table 7.2 The distribution of gender for compact car owners and SUV owners of USA and UK

Country	USA		UK	
Car	Compact car	SUV	Compact car	SUV
Gender				
Men	16	37	8	54
Women	25	12	36	2

Table 7.3 The presence of skin allergies according to life style

	Skin allergy	No skin allergy	Total
Sedentary life style	132	35	167
Nonsedentary life style	17	116	133
Total	149	151	300

Table 7.4 The opinion survey about an agenda A according to the ages

	20s	40s	Total
Agreement	39	78	117
Opposition	61	22	83
Total	100	100	200

Table 7.5 The presence of bronchial disease due to smoke

	Bronchial disease	No bronchial disease	Total
Smoke	90	60	150
Not smoke	60	40	100
Total	150	100	250

Table 7.6 The survey of car color preference by gender

	Purple	Blue	Total
Women	42	73	115
Men	44	71	115
Total	86	144	230

correct answerers) and 6.02 % (13, 10.48 % of the correct answerer) of the students mentioned the concept of conditional probability.

The data in Table 7.7 (item 23) and Table 7.8 (item 25) are correlated but the categories in the predictor variables (row variables) have different sample sizes with one category almost twice of the other. The correct response rates were 37.5 % and 36.11 %, respectively. Also, almost a third of students have chosen “no correlation.” Some of these students who have chosen “no correlation” might have meant they cannot determine the correlation since the reasons they stated were “because the difference in the margin of row is large,” “because the sample size is too small” (for item 25), “because the sample size was different between two groups.”

Table 7.7 The acceptance/reject according to the origin of raw materials

	Accepted goods	Rejected goods	Total
Japanese raw materials	127	58	185
Domestic raw materials	68	15	83
Total	195	73	268

Table 7.8 The preference for imported cars and domestic cars by region

	Imported car	Domestic car	Total
Gangnam-gu	27	24	51
Gangbuk-gu	21	3	24
Total	48	27	75

7.4 Conclusion and Discussion

This study surveyed Korean high school students' understanding of the concept of correlation with a group of items for the data expressed in scatter plots, texts, and contingency tables. As shown by the analysis results in the previous section, the correct response rates were over 70 % in 13 out of 25 items but below 50 % in eight items. Particularly for item 7 and 8, the correct answer rate was below 20 %. From these results, we can derive issues to be discussed and educational implications as follows.

First, students do not have right criteria for determining the strength of correlation presented in scatter plots. Most of students could determine if there is correlation/no correlation and if the correlation is positive/negative by seeing the data presented in scatter plots. However, they seemed not to have more sophisticated criteria to judge about the strength of the correlation. We observed misconceptions such as: (1) if the slope of regression line is steeper, the correlation is stronger; (2) if the data points are relatively close to each other, the correlation is stronger even if the data are same. These misconceptions were demonstrated clearly by the fact that the correct response rate was only 17.13 % for item 7 asking about the slope of regression line and the strength of correlation and 18.98 % for item 8 asking about change in the scale of variables. Both of the two misconceptions arise from the visual characteristics of scatter plots and this is probably because the students had learned only the informal definition of correlation without learning about statistical concepts such as regression line and correlation coefficient.

Second, when data on the strength of correlation were given in text, the students had difficulty in understanding the distribution-related characteristic of the data. Students had difficulty in figuring out the local distribution characteristic of data, which cannot be guessed merely based on the expression "The correlation is strong" without statistical knowledge of correlation. Also, they again have shown confusion between strong correlation and steeper slope. Without clear understanding of the concept of correlation, it will be difficult to interpret a text describing a situation in the real world using the term correlation. Nevertheless, such texts related to correlation

are circulated frequently through various media around us and utilized as ground materials for making various decisions. This suggests the necessity of adequate educational measures for helping students get used to such materials and utilize them in making decisions.

Third, a large number of students did not have the concept of conditional probability or were not able to apply the concept properly to the data in 2-by-2 tables. Most of students did not show clear criteria for judgment of association and they rather approached each problem using intuitive guess comparing absolute count of cases without applying the concept of conditional probability, which is necessary for the decision. Also, we found many students could not determine the presence of correlation even with a contingency table showing the presence of correlation clearly. Some students were hesitant to say “there is correlation” even if the numbers in the table clearly shows strong association between two variables. The reasons for the hesitance were: (1) the sample size is too small; (2) data are the results of random sampling, therefore, not truthful; (3) the sample sizes of two groups are different; (4) they got confused with correlation and causation. These four reasons or misconceptions leading students to wrong intuitive conjectures may be overcome through education of statistical thinking emphasizing decision making through exploratory data analysis.

Summing up the discussion, Korean high school students who had learned elementary knowledge about scatter plot and contingency table showed low-level understanding of the concept of correlation. In particular, this study confirmed that they had misconceptions arising from the visual elements of scatter plots and were unable to interpret texts and contingency tables related to correlation. From these results, we concluded that educational measures are required in order to remove such misconceptions and to improve understanding of correlation. Considering that the current mathematics curriculum does not cover the concept of correlation, we need to improve the curriculum as well.

This study surveyed how the concept of correlation was understood by high school students who had completed the 7th National Curriculum for middle school, but we need to expand/change the survey to include students who had completed the new 2007 Revised National Curriculum, namely, those who had not learned the concept of correlation in secondary school at all. Through in-depth interviews, what is more, further detailed discussions should be made on the causes of misconceptions related to correlation and reasons for students’ difficulty in understanding specific concepts. These research efforts are expected to set the future direction of statistics education in Korea.

References

- Baterno, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27(2), 151–169.
- Cleveland, W. S., Diaconis, P., & McGill, R. (1982). Variables on scatterplots look more highly correlated. *Science*, 216(4), 1138–1141.

- Gal, I. (2004). Statistical literacy. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 227–256). Dordrecht, The Netherlands: Kluwer.
- Ministry of Education, Science and Technology. (2007). *Mathematics curriculum*. Seoul, South Korea: Ministry of Education, Science and Technology.
- 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.
- Zieffler, A. S., & Garfield, J. B. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal*, 8(1), 7–31.

Chapter 8

Connections Between Statistical Thinking and Critical Thinking: A Case Study

Einav Aizikovitsh-Udi, Sebastian Kuntze, and David Clarke

8.1 Introduction

Both Critical Thinking skills and Statistical Literacy are considered as prerequisites for the participation of responsible citizens in democratic societies. Consequently, fostering competencies in these areas is an important goal for schooling internationally. However, research activities in psychology, mathematics education and educational research related to Critical Thinking skills on the one hand and competencies in Statistical Literacy on the other have followed almost separate paths so far—the two domains call for an interconnected perspective. The lack of empirical research appears as astonishing given common foci and intersection areas of both areas seen from a theoretical point of view. Consequently, the project “Connections between Critical Thinking and Statistical Thinking” (CCTST) concentrates on carrying out research aimed at describing how competencies in both of these areas are interconnected and identifying focused and interconnected ways of fostering both of these competencies through classroom instruction. In the first phase, a cross-cultural interview study (Israel, U.S., and Germany) is being carried out in order to generate hypotheses about the interconnectedness of critical and statistical thinking by means of qualitative methods. In this chapter, we present a first part of this ongoing work in the form of a case study. On the basis of evidence collected so far, we have

E. Aizikovitsh-Udi (✉)
Beit Berl College, Kfar Sava, Israel
e-mail: einav.aizikovitsh@beitberl.ac.il

S. Kuntze
Ludwigsburg University of Education, Ludwigsburg, Germany
e-mail: kuntze@ph-ludwigsburg.de

D. Clarke
University of Melbourne, Melbourne, VIC, Australia
e-mail: d.clarke@unimelb.edu.au

tentatively identified aspects of critical thinking with possible links to statistical thinking and vice versa.

The first section of this chapter reviews key elements of the theoretical background, relating statistical thinking to critical thinking. The second section introduces the research questions. Information about the research methodology is given in the third section of this chapter, and results are reported in the fourth section. The fifth section contains a short discussion of the evidence and conclusions.

8.2 Theoretical Background

8.2.1 *Statistical Thinking and Critical Thinking*

In a well-known definition of Statistical Literacy by Gal (2004), a “critical stance” is included among the key attitudes for successful statistical thinking (ST)—hence, Gal includes such attitudes in the notion of statistical literacy (cf. also Wallman, 1993; Watson, 1997; Reading, 2002). However, being critical in statistical contexts is not only an attitude, but it is possible to describe specific abilities that have to be used in order to critically evaluate statistical data. Two key concepts or overarching ideas in statistical thinking relevant for a critical evaluation of data are *manipulation of data by reduction* (Kröpfel, Peschek, & Schneider, 2000) and *dealing with statistical variation* (e.g. Watson & Callingham, 2003). Successfully manipulating data by reduction requires the awareness that, e.g. calculating a mean value affords an overview on the original data, but it reduces the initial information. Hence, the resultant statistical value is (only) an indicator corresponding to a specific mathematical model, and we should not forget that it reflects only a part of the information. For critically evaluating the data, we might need additional information about the distribution, such as the variance, or information about extreme values. Dealing successfully with statistical variation means that an awareness of chance variability is necessary when evaluating data (Wild & Pfannkuch, 1999). A value in statistics often cannot be taken for granted as being exact, it might be different, e.g. as a consequence of a second survey. Such knowledge about statistical variation and abilities for dealing with it can be described in terms of critically evaluating data, questioning its stability and an awareness of the significance of differences. Another specific example consists in the aspect of *sampling*. Watson, Kelly, Callingham, and Shaughnessy (2003) describe abilities related to sampling as a sign of a deeper understanding of statistics and hence of advanced elements of ST. Indeed, the awareness of where the data comes from, of the sample size and of the choice of the sample is crucial to the quality of any statistical analyses—and it can again be seen as an ability related to critical thinking, as, for example, inductive reasoning as well as questioning evidence is involved.

Critical thinking (CT) skills rely on self-regulation of the thinking processes, construction of meaning and detection of patterns in supposedly disorganized structures. Critical thinking tends to be complex and often terminates in multiple solutions that

have advantages and disadvantages, rather than a single clear solution. It requires the use of multiple, sometimes mutually contradictory criteria, and frequently concludes with uncertainty. This description of the notion of CT already suggests links with ST, such as dealing with uncertainty, contradictions and a critical evaluation of given claims (cf. McPeck, 1981). Dealing critically with such information—a crucial aspect for both domains—demands critical/evaluative thinking based on rational thinking processes and decisions (Aizikovitsh-Udi, 2012; Aizikovitsh-Udi & Amit, 2008).

The two key concepts of manipulation of data by reduction and dealing with statistical variation have been implemented in a corresponding hierarchical competency model of a sub-aspect of statistical literacy (Kuntze, Lindmeier, & Reiss, 2008; Kuntze, Engel, Martignon, & Gundlach, 2010), using the core metaphor of data-related reading (cf. Curcio, 1987). According to the considerations earlier we can expect that the competency of “using models and representations in statistical contexts” encompasses aspects of critical thinking. These theoretical connections seem very evident, but can they be distinguished empirically in thinking processes when solving tasks related to Statistical Thinking or Critical Thinking?

8.2.2 Empirical Findings Relating Critical Thinking to Statistical Thinking

There are prior studies in which relationships between CT and ST have been investigated in a quantitative correlational design. For instance, Royalty (1995) found a correlation of $r = .49$ between the scores on the Cornell CT Test (Ennis & Millman, 1985) and on a selection of statistics items. However, Royalty (1995) does hardly describe the structure or the model associated with this ST instrument, nor does he discuss the correlation he found from a content point of view. His study with 109 participants rather focuses on the question of generalizability of CT, and ST is considered as a domain-specific variable. This research calls for in-depth analyses about *how* CT and ST may interdepend, and the findings of Royalty indicate that such in-depth analyses may yield interesting and important possible explanations of the correlations which open up the way also for follow-up quantitative research.

8.3 Research Questions

Consequently, this study aims to provide evidence for the following research question: How is critical thinking connected with statistical thinking? How can the simultaneous interpretation, using both Critical Thinking and Statistical Thinking, of an individual’s problem solving help to better explain evidence related to thinking processes in either domain? Beyond the theoretical character of this research interest, these questions are also relevant for the context of seeking to inform classroom instruction.

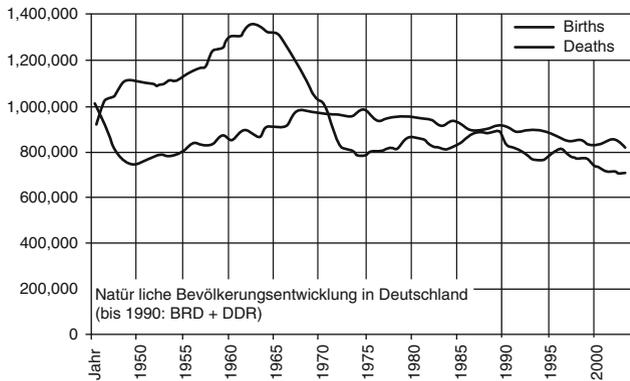
8.4 Methods

For exploring thinking processes related to tasks in the domains of both Statistical Thinking and Critical Thinking, individual semi-structured interviews are being conducted with teachers. The interviews focus on thinking aloud when solving tasks and last about 40–50 min. Beyond solving the tasks, the interviewed persons are also free to give their personal view on the tasks, respectively. We report in the following section results of one interview. The bottom-up analysis of the evidence that has been carried out concentrated on thinking processes relevant from the point of view of Statistical Thinking (ST), on the one hand, and of Critical Thinking (CT), on the other. For attaining this goal, a first analysis of the evidence was done focusing on an interpretation against the background of ST only, then a second analysis concentrated on an interpretation employing a CT point of view. The analysis was carried out in a bottom-up-approach, focusing on criteria from the theoretical frameworks, respectively (Ennis & Millman, 1985; Watson, 1997). In a third step, we carried out a combined interpretative analysis, drawing on both approaches, in order to explain thinking processes in more depth and to highlight relationships between CT and ST elements. In this methodological approach, the analyses were done by two raters.

8.5 Results

Our aim in this study was to exercise and evaluate CCTST. In order to answer the research questions, the findings from this study were analysed from the perspectives of Critical Thinking and Statistical Thinking separately and then by a joint perspective, discussing possible connections between the two. We would like to illustrate this with a first example taken from the interview with Nena, a U.S. secondary teacher. Nena was an experienced Mathematics instructor with 20 years of mathematics teaching experience. She had been teaching in the same school for 12 years. According to observations by the first author, her teaching is very detailed and precise and is accompanied by oral and written explanations. Nena gives many instructions during the process of solution: how to solve equations, perform calculations, work with models, when to contract, etc. Nena solves each equation until the final result and does not skip stages in problem simplification, including detailed substitutions and all the necessary calculations. Her teaching is best described as “traditional” and extensive descriptions of such teaching can be found in the literature documenting traditional teaching (e.g. Chazan, 2000). Nena is dedicated to improving her teaching wherever possible and had been participating in a year-long professional development programme for mathematics teachers at the time this interview occurred. In the interview, Nena had been asked to solve the problem in Fig. 8.1, among others, while thinking aloud. Here is the corresponding part of the interview with Nena:

„The Germans don't have enough children”, was a recent newspaper's headline. Consider the following diagram about the development of the German population:



From which year on has the population been decreasing?

Fig. 8.1 Task “population” (Kuntze et al., 2008)

Interviewer: What do you think? From which year on has the population been decreasing?

Nena: Well... the population has been decreasing since 1963 until about 1973, where the population begins to rise a bit... but it is still down since 1963....

Interviewer: Are you confident about it?

Nena: of course... I am sure! You can look at this problem mathematically, anywhere there is a positive slope you could say the population is increasing, where there is a negative slope the population is decreasing. But compared to 1963, it's always been down. Am I right?

Interviewer: Sorry, I can't say.

Seen from the perspective of Critical Thinking (CT), the analysis yielded that in this part of the interview, dealing with assumptions is one of the key elements (Ennis & Millman, 1985). Nena initially has an assumption, interpreting the graph of births as completely determining the population development. Even when asked to reflect on this assumption, Nena does not generate possible counter-arguments for testing her initial assumption, nor does she appear to question this assumption. Consequently, her way of dealing with assumptions appears to be unsuccessful; she tends to seek for confirming evidence rather than for evidence that might challenge her initial assumption.

Seen from the perspective of Statistical Thinking (ST), Nena chooses an inappropriate statistical model for interpreting the data given in the diagram (cf. description of this task in Kuntze et al., 2008). She appears to focus on the data related to the births only, and she deduces her conclusions from a mathematical consideration of slopes. Even when encouraged by the interviewer, she does not check this model against the background of the full data given in the diagram.

Consequently, the (separate) analyses suggest that Nena's answers show deficits both in Critical and Statistical Thinking. Looking at the relationships between CT and ST in more depth, we attempt a combined analysis: At the very beginning of Nena's thinking process, she shows a partial or incomplete perception of the evidence, focusing on the birth data from 1963 on. This selective focus may have been a result of the headline given in the diagram ("The Germans don't have enough children"), which can be interpreted as an assumption in CT terms. This headline may hold from about 1972 on, or, within Nena's misinterpretation of the births determining the population, from the turning point in 1963 on. Nena does not question the given headline, which would have been a characteristic of CT. This stance of questioning assumptions derived from data is also an important aspect of ST performance. In the next statements by Nena, it becomes even more apparent just how important questioning assumptions can be for ST: Nena does not question her first—and partial—interpretation of the diagram and hence does not challenge her model of the data in terms of ST, neither by looking at the complete curve of the births, nor by developing a model of considering the population as a sort of "reservoir" with "incomings" and "outgoings". The latter would permit to discover that the intersection point(s) of the curves are meaningful. In this context, it is interesting that Nena emphasises that it is possible to "look at this problem mathematically", which suggests that she sees a discrepancy between looking at the situation from the perspective of a mathematical model and looking at it from the perspective of the context (population and children born in Germany). Possibly the mathematical or statistical model is considered as an authority that is used to justify the appropriateness of the assumption instead of questioning the model chosen initially. We may conclude from this combined analysis that the shortcomings in CT go hand in hand with shortcomings in ST and that the elements of reasoning in both domains interfere and interact. Hence, help in either domain, i.e. both in CT and ST, could possibly have had a positive impact on the thinking process as a whole. Moreover, the CT and ST approaches can give not only simultaneous and parallel ways of interpreting the reasoning process, but, through a combined analysis, can explain even better how knowledge and awareness of both CT and ST can be mutually beneficial, reinforcing related reasoning approaches. However, CT and ST are not always interdependent in an obvious way, as the following example suggests (cf. Fig. 8.2 and the corresponding interview section): Aizikovitsh-Udi, Kuntze and Clarke

Interviewer: So, what do you think, with which of the following statements do you agree?

Nena: I would agree with the third statement.

Mrs. Blum would like to buy a reliable Laptop, either a C-Pad or an S-Top. In a computer magazine, 400 laptops of each brand have been tested. In this comparison the C-Pad has turned out to be more reliable. In the evening she talks to three friends. Two have S-Tops and never had problems. The third had a C-Pad, but had so many hardware problems with it, that he has sold it again immediately.

With which of the following statements do you agree?

- Mrs. Blum should buy an S-Top, because the friend with the C-Pad had made bad experiences, whereas the friends are happy with the S-Tops.
- Mrs. Blum should buy a C-Pad, because the test in the computer magazine is based on a high number of computers, not only on one or two.
- No matter how she decides, it can happen that she gets a Laptop that causes problems frequently.

Fig. 8.2 Task “Laptops”

Interviewer: Can you explain, please?

Nena: Yes ... 400 computers is not a large sample when talking about computers so I would go with the third statement and just listening to the comments of her friends and the consumer’s magazine, it is possible that both the computers could be just so.

Interviewer: Are you sure?

Nena: Sure. I like this question!

Seen from the perspective of CT, an important aspect of Nena’s answers is that she questions not only the experiences of the friends, but also the results from the study with the 400 laptops. On this base, she expresses agreement only with the third statement, which corresponds to a high level of questioning evidence as a sub-aspect of CT.

However, considering Nena’s answers under the lens of ST, she appears not to fully acknowledge the statistical power of the sample of 400 laptops. So even though Nena remarks that it is not possible to make a prediction on the base of the data, and even if she appears to compare the number of the 400 laptops to the number of all laptops, she does not reflect in depth on the higher statistical power of the computer magazine study.

In a joint perspective, including aspects from CT and ST, Nena might even have had to face a conflict: Her CT evaluation of the situation may have lead to a view that predictions are impossible and that a laptop of any type may cause problems. This dominant critical attitude may have somewhat blocked the awareness of elements of ST, e.g. reasoning related to the sample size and representativeness. This section of the interview may at first sight suggest that CT and ST were relatively independent here, because a high level of CT coincides with incomplete ST processes. However, as CT and ST have probably interacted in Nena’s thinking process, with CT “blocking” a deeper ST analysis, this example also gives insight how CT and ST may interfere. An increased awareness of ST could even have contributed to the development of CT in this example: e.g. when a decision for a laptop has to be made, acknowledging the study with the 400 laptops can be seen as a form

The police president shows the following diagram and says: "This diagram shows, that since 2005, the number of crimes in the city center has increased, so that we have to expect that it will further go up in the next years. We need more policemen for patrol in the city center."

In a more extensive press communication, Fred has found the data of the last 7 years:

Year	2000	2001	2002	2003	2004	2005	2006
Number of crimes	504	528	525	499	529	518	538

Do you agree with the interpretation of the police president?
Why or why not?

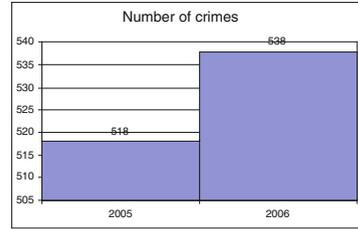


Fig. 8.3 Task “crimes”

of thinking critically about one’s own critical thinking, i.e. questioning the initial assertion that predictions are impossible.

Conversely, also ST can be dominant over CT, as the following example (associated to the problem in Fig. 8.3) suggests:

- Interviewer: ...and here... Do you agree with the interpretation of the police president? Why or why not?
- Nena: I do not see a big difference with the number of crimes for any given year that would warrant extra police force to be hired....
- Interviewer: Can you justify it, please?
- Nena: It looks like the average is approximately 520 which is close to all the numbers so I do not think anything different is really happening from any given year.

According to Ennis’ taxonomy (Ennis & Millman, 1985) one crucial element of CT is raising questions, having doubts and exploring the definition of important notions like “crime”. In our case it is very important to know about the nature of the crimes for drawing conclusions. For example, if we knew that all the crimes were murders, we would decide differently than if these crimes were related to paying taxes or fraud. In our case, no question about such a definition was raised and the focus was just on the pure numerical aspect. Hence, in terms of CT, Nena doesn’t show the complete spectrum of CT skills.

From the point of view of ST, Nena uses an appropriate model and shows an awareness of statistical variation. By these means, she arrives at the conclusion, that, given the variation of the data of the past years, the rise of the crime number is not significant. Consequently, seen from the ST perspective, Nena shows an appropriate understanding of the statistical situation.

Looking at this part of the interview in the joint CT and ST interpretation mode, the analysis yields that Nena successfully questions the statement of the police president by using the data given in the problem and a statistical argumentation. In the following, she appears to remain in the statistical domain, giving more details related to the model she had chosen (distances to the average value). This focus on the ST model may have hindered her use of any of the CT skills in the following, e.g. analysing and questioning the definition of “crime” as the key notion here, questioning the evidence (i.e. the way the data had been collected), etc.

A company produces two sorts of headache tablets. Both sorts have been tested in a laboratory with respectively 100 persons suffering from headache. The diagram below shows, how long it took until the headache was over. Each point represents one test person.

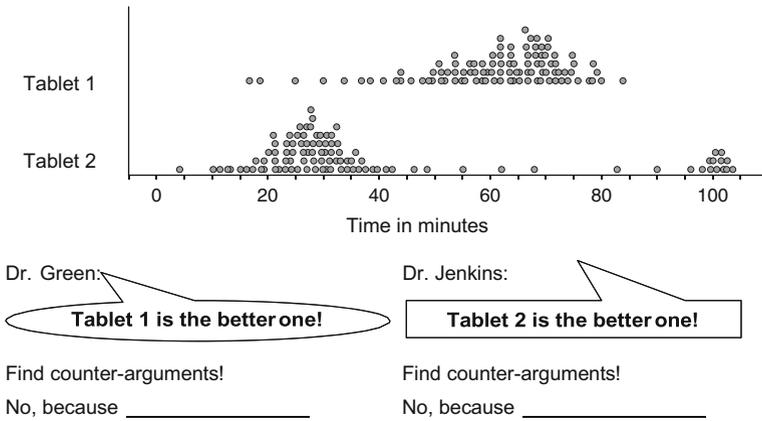


Fig. 8.4 Task “tablets” (Kuntze et al., 2008)

Questioning data plays a role also in the following example related to the problem in Fig. 8.4:

Nena: And here ... I think tablet one takes too long to get rid of the headache. Tablet two seems to get rid of the headache a lot quicker for the majority of the people. You don't really know the age or weight of the people. Many factors play into the reason why a headache might occur so the statistics are poor. Based on the chart, I would have to pick tablet number two in hope of a speedy recovery.

From the perspective of CT, Nena not only evaluates the given statements, but she also shows CT elements when going beyond the data given: She gives examples of relevant influencing factors, and questions the data provided in the diagram (“the statistics are poor”).

From the point of view of ST, the analysis of Nena’s short answer yields that Nena chose an appropriate model and was aware of the key elements of the problem, even if she did not explicitly discuss the minority of cases with very slow recovery for tablet 2. These considerations led to her personal conclusion to pick tablet number two, as she obviously sees the chance of a “speedy recovery” as more important than the risk of a very slow recovery.

Looking at both CT and ST, the example appears to highlight how elements of CT can contribute to ST, e.g. when evaluating data, its presentation and analysis, planning data collection, etc. In the example, Nena suggests an analysis that takes into account the age or weight of the persons in the study. Conversely, aspects of ST

like dealing with statistical variation and uncertainty can contribute to CT, especially when it comes to decisions in non-determinist situations, where full data is unavailable.

8.6 Discussion and Conclusions

This exploratory study investigates CCTST and demonstrates that these connections clearly exist at the level of individual reasoning practices. Specific CCTST are suggested by the evidence related to the thinking processes elicited in the interview. It might be said that an individual employing Statistical Thinking had access to a structured framework of analytical principles that could guide and support their reasoning. That is, the relationship between measures of central tendency and variance, for example, structure any consideration of distribution of data that might be invoked in drawing evidence-based conclusions or making evidence-based judgments. On the other hand, the components of Critical Thinking are not related in such a structured fashion and an individual's inclination to employ one strategy (e.g. Questioning Evidence or Questioning Assumptions) can be given expression without any obligation to also invoke other components of Critical Thinking. Some Critical Thinking skills resemble the "heuristics" that were the focus of the enthusiasm for problem solving in the 1980s and 1990s (Clarke, Goos, & Morony, 2007). Catalogues of such heuristics were similarly fragmented. Ennis and others have catalogued critical thinking skills (Ennis & Millman, 1985) and even arranged these categories in a form of hierarchy, but the connection between specific critical thinking skills is under-theorised in comparison with Statistical Thinking. Nonetheless, the forms of Critical Thinking identified in such classificatory schemes are clearly of significance, both as aspects of reasoning and as potential curriculum content. The question of how best to conceptualise these skills, how to integrate or connect them with other curricular goals and how best to promote them and nurture their development in the classroom remains a major challenge. The case study presented in this chapter has made some aspects of that challenge explicit.

When the individual being offered the opportunity to employ Statistical Thinking and/or Critical Thinking is a mathematics teacher (Nena), with more or less complete access to the principles and practices of Statistical Thinking, then it seems reasonable to anticipate the dominance of Statistical Thinking as a characteristic of the individual's reasoning processes. In fact, Nena's inclination to question evidence served as a general behavioural reference point and was capable of overruling statistical considerations, while avoiding any obligation to employ complementary Critical Thinking skills such as attempting to generate counter examples or identifying and questioning assumptions. The lack of an established theoretical structure in Critical Thinking seriously restricts the utilization of Critical Thinking in the manner in which Statistical Thinking can be employed. However, if it were possible to develop a structure for Critical Thinking in which the component elements were not only identified, but also their relationship established, then to invoke one aspect

of Critical Thinking would serve to catalyse the use of other related aspects, because the connections between elements would be well known and understood.

Since it has been (1) argued, and (2) demonstrated in this chapter that ST requires many of the reasoning processes essential to CT, it is possible that careful education in the use of ST might be employed as the entry point for specific instruction in CT. An earlier study (Aizikovitsh-Udi, 2012) has documented efforts to produce CT through a programme of instructional immersion in the related topic of probability. The arguments and data reported here suggest that similar efforts might be made to develop a structured programme of instruction in ST that also integrated analogous elements of CT. In such a programme, the structure and utility of CT would be elaborated by analogy with related aspects of statistical thinking requiring reasoning processes common to both ST and CT. This proposal warrants investigation.

The case study reported in this chapter suggests several imperatives if such a programme is to be realised: (1) A structure must be found for Critical Thinking that consists of more than just a list of components (hierarchical or not); (2) aspects of Statistical Thinking must be identified that are analogous to corresponding aspects of Critical Thinking; (3) an instructional programme must be devised that provides the opportunity to employ Statistical Thinking, while simultaneously introducing students to the practices and structure of Critical Thinking; and (4) teacher education programmes must initiate pre-service (and in-service) teachers into not only the instructional practices required for such a programme but the thinking skills themselves (characteristic of both Statistical Thinking and Critical Thinking).

References

- Aizikovitsh-Udi, E. (2012). *Developing critical thinking through probability models, intuitive judgments and decision-making under uncertainty (Doctoral dissertation)*. Saarbrücken, Germany: LAP Lambert. ISBN 978-3-8383-7240-2.
- Aizikovitsh-Udi, E., & Amit, M. (2008). Developing critical thinking in probability lesson. In O. Figueras, J. L. Cortina, S. Alatorre, T. Rojano, & A. Sepúlveda (Eds.), *Proceedings of the Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 9–13). Morelia, México: Universidad Michoacana de San Nicolás de Hidalgo.
- Chazan, D. (2000). *Beyond Formulas in Mathematics and Teaching: Dynamics of the High School Algebra Classroom*. New York: Teachers College Press.
- Clarke, D. J., Goos, M., & Morony, W. (2007). Problem solving and working mathematically: An Australian perspective. *Zentralblatt für Didaktik der Mathematik*, 39, 475–490.
- Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18(5), 382–393.
- Ennis, R. H., & Millman, J. (1985). *Cornell critical thinking test—Level Z*. Pacific Grove, CA: Midwest.
- Gal, I. (2004). Statistical literacy, meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht, The Netherlands: Kluwer.
- Kröpfl, B., Peschek, W., & Schneider, E. (2000). Stochastik in der Schule: Globale Ideen, lokale Bedeutungen, zentrale Tätigkeiten. [Stochastics in school: Global ideas, local significances, central activities]. *Mathematica Didactica*, 23(2), 25–57.

- Kuntze, S., Engel, J., Martignon, L. & Gundlach, M. (2010). Aspects of statistical literacy between competency measures and indicators for conceptual knowledge—empirical research in the project RIKO-STAT. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of ICOTS 8*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from www.stat.auckland.ac.nz/~iase/publications.php.
- Kuntze, S., Lindmeier, A. & Reiss, K. (2008). “Using models and representations in statistical contexts” as a sub-competency of statistical literacy—results from three empirical studies. *Proceedings of the 11th International Congress on Mathematical Education (ICME 11)*. Monterrey, Mexico: Universidad Autonoma de Nuevo Leon. Retrieved from <http://tsg.icme11.org/document/get/474>.
- McPeck, J. (1981). *Critical thinking and education*. New York: St. Martin’s Press.
- Reading, C. (2002). Profile for statistical understanding. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved January 28, 2008, from http://www.stat.auckland.ac.nz/~iase/publications/1/1a4_read.pdf.
- Royalty, J. (1995). The Generalizability of Critical Thinking: Paranormal Beliefs Versus Statistical Reasoning. *The Journal of Genetic Psychologist*, 156(4), 477–488.
- Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1–8.
- Watson, J. M. (1997). Assessing statistical thinking using the media. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107–121). Amsterdam, The Netherlands: Ios Press.
- Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3–46.
- Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students’ understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, 34(1), 1–29.
- Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 3, 223–266.

Chapter 9

Tasks Associated to the Treatment of Tables at Elementary School and Its Level of Difficulty

Soledad Estrella, Arturo Mena-Lorca, and Raimundo Olfos

Piaget and Inhelder (1955) argued that some types of tables should be presented to students who are at least 12 years old. In turn, Duval (2003) argued that various types of tables and their interpretations demand different cognitive resources, and that it is relevant to explain that not all tables are the same, and not all students understand in the same way; moreover, students' difficulties with tables can endanger other learning. What concepts are present in students' mind during their interpretation of tables? What are the main semiotic conflicts in the task? In 2010, two Chilean studies, (Estrella, 2010; Estrella, Olfos & Mena-Lorca, 2015) referred to pedagogical content knowledge, showed that treatment of tables presents difficulties for teachers ($n=85$) and students ($n=1500$). This research analyses various tasks associated to the treatment of tables at elementary school and their level of difficulty, according to Curcio's taxonomy of graphical understanding (Curcio, 1989; Friel, Curcio, & Bright, 2001).

Some difficulties that would arise in the treatment of the tables: reading headers and titles is complex because it differs from the verbal register by eliminating the syntactic organization of sentences; it can be read like the graphics or plans, as available location goes from vertical/horizontal, or diagonal, or both; cognitive demand to complete a table inside (body of data) is smaller than to construct margins of the table; there are different levels of reading comprehension of the table by number of variables in play to categorize (based on a variable, according to a second variable, and according to both variables); there are different levels in the

S. Estrella (✉) • A. Mena-Lorca • R. Olfos
Universidad Católica de Valparaíso, Valparaíso, Chile
e-mail: soledad.estrella@pucv.cl; arturo.mena@pucv.cl; raimundo.olfos@pucv.cl

construction of the table by number of variables in play to categorize (depending on a variable, according to a second variable, and as the two variables); there are also difficulties in counting the items in each cross-category because the cross-classification may be a necessary but not sufficient condition to construct a table.

References

- Curcio, F. R. (1989). *Developing graph comprehension*. Reston, VA: N.C.T.M.
- Duval, R. (2003). Comment analyser le fonctionnement représentationnel des tableaux et leur diversité? *Spirale*, 32, 7–31.
- Estrella, S. (2010). *Instrumento para la evaluación del conocimiento pedagógico del contenido de estadística en profesores de primaria*. (Unpublished Master's Thesis). Pontificia Universidad Católica de Valparaíso: Chile.
- Estrella, S., Olfos, R., & Mena-Lorca, A. (2015). Pedagogical content knowledge of statistics among primary school teachers. *Revista Educacao e Pesquisa*, 41(2), 477–493.
- Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124–158.
- Piaget, J., & Inhelder, B. (1955). *De la logique de l'enfant à la logique de l'adolescent*. Paris: Puf.

Chapter 10

An Analysis of the Statistical Contents Covered In China, Singapore and Taiwan Mathematics Textbooks at the Primary Level

Hak Ping Tam and Ou Yung Chih

10.1 Introduction

Items on probability and statistics account for a sizable percentage in both TIMSS and PISA. Yet their coverage in Taiwan's mathematic curricula has been gaining ground rather slowly. Moreover, the statistical content is presented more as a set of individual techniques rather than organized around some big ideas. In order to vitalize the statistic curricula in Taiwan, an attempt is initiated to compare local primary mathematics textbooks with those from China and Singapore in relation to how they differ in their arrangement of statistic. It is believed that useful insight can be gleaned via a careful comparison across all versions of textbooks.

10.2 Methodology

All three available versions of Taiwan textbook are included in this study. The Beijing Normal University's version and *My Pals Are Here* are chosen to represent textbooks from China and Singapore, respectively. Every version carries twelve textbooks at the primary level. Comparisons are made regarding the topics being covered, their coverage, grade levels in which they are introduced, their sequences of presentation as well as the way the concepts are discussed. The analysis was done in 2012.

H.P. Tam (✉) • O.Y. Chih
National Taiwan Normal University, Taipei, Taiwan
e-mail: t45003@ntnu.edu.tw

10.3 Results

It was found that the three Taiwan versions have a relatively lower coverage on statistics content. They tend to be introduced at grades four, five and six, rather than spreading out throughout all primary grades as in their counterparts. Beijing Normal's version covered more topics and at greater length than all the other versions. It contained more review materials to refresh students' memory of concepts learned previously.

Chapter 11

Mathematical Modelling for Critical Statistics Education

Luana Oliveira Sampaio and Maria Lucia Lorenzetti Wodewotzki

This study involved 55 students of the discipline Statistics for Administration. With them, we conducted an extracurricular activity through Mathematical Modelling (MM) in an attempt to create conditions for a Critical Statistics Education (CSE). In this line, MM involves investigations of reality, while providing an environment where students gain space for discussions. Those discussions that have some connection with the construction of the Mathematical Model is what Barbosa (2006) calls Modelling Routes, which consists of three types of discussions: mathematical, technological and reflexive. However, outside the Modelling Routes, there are the Parallel Discussions, that refers to general aspects of the problem without being used in the construction of the mathematical model. In our research we identified new branches within the Parallel Discussions (SAMPAIO, 2010): Mathematical Parallel Discussions: refers to the speeches belonging to the field of pure mathematics and/or statistics; Technological Parallel Discussions: refers to the way, or set of procedures and tools used to plan, develop and evaluate the learning environment; Reflexive Parallel Discussions: refers to ideas related to aspects of social life and can involve socio-critical interpretations of the results of mathematical or statistical studies; Other Parallel Discussions: refers to the speeches that do not meet the definitions above. Faced with this classification, we observe that the discussions outside the Modelling Routes can also have an important role in MM Practices. We conclude that the Reflexive Discussions and Reflexive Parallel Discussions contribute more to the development of CSE, that, according to Campos (2007), brings together the goals of Statistics Education and goals of Critical Education in order to produce a pedagogy that is democratic, reflective and engaged in its greatest role of social responsibility with students.

L.O. Sampaio (✉) • M.L.L. Wodewotzki
Luana Oliveira Sampaio UFSB (Federal University of Southern Bahia), Maria Lucia Lorenzetti Wodewotzki (São Paulo State University), Rio Claro, Brazil
e-mail: luanaoliveirasampaio@gmail.com; mariallw@rc.unesp.br

References

- Barbosa, J. C. (2006). *Students' discussion in mathematical modelling*. Paper presented at 3rd International Conference on the Teaching of Mathematics. Istanbul, Turkey.
- Campos, C. R. (2007). *A educação estatística: uma investigação acerca dos aspectos relevantes à didática da estatística em cursos de graduação*. Tese (Doutorado)—IGCE, Universidade Estadual Paulista, Rio Claro, Brasil.
- Sampaio, L. O. (2010). *Educação Estatística Crítica: uma possibilidade?* Dissertação (Mestrado)—IGCE, Universidade Estadual Paulista, Rio Claro, Brasil.

Chapter 12

Survey and Research on the Levels of High School Students' Critical Evaluation of Statistical Information and the Influence Factors

Ma Ping

12.1 Study Questions

With the rapid development of economic globalization and information society, statistics, as an important tool, is widely used in many fields. According to Gal (2002), in order to satisfy the needs of the society, we should have “the ability to explain and give critical evaluation on statistical information, viewpoints relating to data, and random phenomena, under different situations.” According to Watson and Moritz (1997) from Australia, there are three levels in the understanding of statistical conception, and the highest level is “to be able to observantly raise questions for improper use of statistics.” (The other two levels are “to understand basic statistical conception” and “to understand statistical language and concepts used in various backgrounds.”) Both Gal and Watson emphasize that critical evaluation of statistical information is one of the important sides of statistical literacy. One of the important objectives for statistical teaching in many countries is to teach students to use statistical knowledge to carry out critical analysis and evaluation on statistical statements.

In China, with the publication of Curriculum Standards in the Phase of Full-time Compulsory Education in 2011 and the Mathematics Curriculum Standard for Senior High School in 2003, Statistics, together with “number and algebra” and “space and graphics,” has become an important part in mathematics teaching. For the first time, in the Curriculum Standard, “statistical conception” becomes one of the important objectives in statistical teaching. According to Dan, Jiasheng, and Chunli (2002), “statistical conception” includes not only the consciousness of thinking questions from the point of view of statistics but also the process of collecting, describing, and analyzing data by oneself and then making appropriate

M. Ping (✉)

The High School Affiliated to Renmin University, No. 3, Unit 2, Shuangyushu Nanli,
Haidian District, Beijing
e-mail: mmp91@163.com

judgment. Therefore, in the implementation of the new Curriculum Standards, how is the level of critical evaluation of statistical information for Grade One students in senior high school? What are the influence factors? How does it influence the development of critical thinking to let students participate in a statistical activity? How should mathematics teacher cultivate students' critical thinking ability in statistical teaching in senior high school? This study focuses on these questions.

12.2 Methodology

12.2.1 Sample Selection

The sample is selected from a key senior high school in Jinan, Shandong Province, China, which includes 229 Grade One students. There are two phases in the study. In the first phase, 134 students were organized into several teams to carry out a statistical activity which they are interested in, and the results (including statistical report, statistical poster, and PPT) were exhibited in the school. In the second phase, 229 students, including 122 girls and 107 boys, received tests based on the questionnaire. The gender ratio is more or less balanced, so as to analyze if there is gender difference in critical evaluation of statistical information. One hundred and seven students participated in the statistical investigation and 112 students did not, so as to analyze the influence of statistical investigation on critical evaluation of statistical information.

12.2.2 Questionnaire Preparation

The questionnaire includes four questions. Two of them are from the study of Kazuhiro (2007), and the other two were prepared by the author and some other mathematics teachers of the high school. Three principles are followed in the selection and preparation of the test questions: (1) All questions require students to give explanation for their answers, so as to distinguish students' level of critical evaluation of statistical information. (2) The situations of the test questions refer to the division of situations in PISA test (OECD 2006). Considering the "distance" from the students' daily life, four situations are selected, namely, personal situation, educational situation, social situation, and scientific situation, so as to analyze the relation between situations with different "distance" and the students' level of critical evaluation of statistical information. (3) Statistical diagrams in the test questions include column diagram, diagram with a single variable, bar diagram, data point diagram with various variables, and scatter diagram, so as to study the influence of different types of statistical diagrams on the students' level of critical evaluation of statistical information.

For each question, the statistical diagrams are provided first and a statistical judgment is given. Then the students are required to judge if the statistical judgment is reasonable or not, and to give their own reasons.

Before the formal test, the author carried out a pretest in another class in the same Grade. Based on the results of the pretest, the author made some adjustment for the order of some questions and modified the wording of some questions. Two days later, the formal test was carried out.

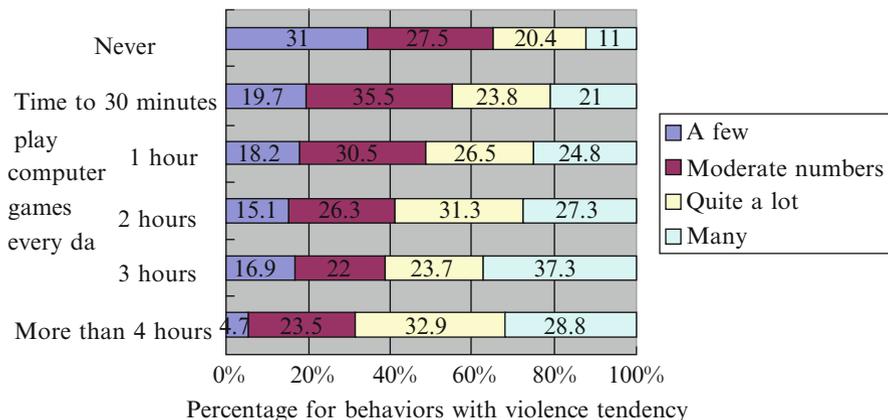
12.2.3 Classification Standard for Levels of Critical Thinking

The question analysis method in this study refers to the study of Kazuhiro (2007). Based on SOLO classification method (Biggs & Collis, 1982) (prestructure level, single structure level, correlation level, and extended abstract level), and combined with Curcio's (1987) three levels of statistical diagrams reading (data reading, reading of the information between data, and reading of the information above data), the classification standard for students' levels of critical evaluation of statistical information (see Table 12.1) is specified according to the relations between data, reasoning process, and explanation above statistical diagrams. Question 2 is used as an example to explain classification standard for levels of critical thinking.

Question 2: The following diagram is the result of a survey for primary school students. The vertical axis shows the time to play computer games every day, and the horizontal axis shows the chance for the occurrence of behaviors with violence tendency (e.g., push or kick classmates, pull other students' hair). (From left to right, there are "a few," "moderate numbers," "quite a lot," and "many.")

Based on the diagram, someone concludes that the reason for the occurrence of violent behaviors among primary school students is that they play computer games for quite a long period of time. Do you agree? Please give your reason.

- A. Agree B. Not agree C. Neither A nor B D. I don't know



In this study, the classification standard for students' levels of critical evaluation of statistical information is as follows (see Table 12.1):

Based on the standard, the author classified the students' answers to each question into five levels, from 0 to 4. Then SPSS was used to analyze the data, and the results are as follows.

Table 12.1 Classification standard for students' levels of critical evaluation of statistical information

Level	Classification standard	Students' answers (take question 2 as an example)
4	<ul style="list-style-type: none"> Being able to find out the relations between data in statistical diagrams; 	<p>According to the diagram, we can find out that playing computer games for a long period of time may cause more violent behaviors. However, playing computer games is not the cause of violent behaviors. The conclusion is not complete. There are many other factors that may cause violent behaviors, such as human relations, characters, family environment, etc.</p>
	<ul style="list-style-type: none"> Being able to make correct reasoning and explanation to the rationality of the statistical statement, according to the relations between data; 	
	<ul style="list-style-type: none"> Being able to give other reasons besides the information of statistical diagrams 	
3	<ul style="list-style-type: none"> Being able to find out the relations between data in statistical diagrams; 	<p>According to the diagram, we can find out that the occurrence of violent behaviors is related to time for playing computer games. However, playing computer games is not the cause of violent behaviors. Playing computer games may have great influence on violent behaviors</p>
	<ul style="list-style-type: none"> Being able to make correct reasoning and explanation to the rationality of the statistical statement, according to the relations between data; 	
	<ul style="list-style-type: none"> Not being able to give other reasons besides the information of statistical diagrams 	
2	<ul style="list-style-type: none"> Being able to find out the relations between data in statistical diagrams; 	<p>According to the diagram, the level of students' violent behaviors increases with the increase of time to play computer games, with slight fluctuation. Therefore, playing computer games for a long period of time is the cause of violent behaviors</p>
	<ul style="list-style-type: none"> Not being able to make correct reasoning and explanation to the rationality of the statistical statement, according to the relations between data 	
1	<ul style="list-style-type: none"> Being able to find out the relations between data in statistical diagrams; 	<p>The students only choose A, B, or C, without giving explanation</p>
	<ul style="list-style-type: none"> Not being able to provide reasoning process 	
0	<ul style="list-style-type: none"> Not being able to find out the relations between data in statistical diagrams (choosing "I don't know," not answering the question or making irrelevant answers) 	<p>I play computer games everyday, but never fight with other people</p>
		<p>Indulging oneself with fictitious computer games for a long period of time may cause people to have hallucination, etc.</p>

12.3 Conclusion and Analysis

12.3.1 *The Students' Overall Level of Critical Evaluation of Statistical Information*

Table 12.2 shows that the critical thinking levels of the students who received the test are distributed in an “olive” shape. The number of 0 and 4 levels account, respectively, for about 10 % of the total number of the students. A total of 62.5 % of the students (reaching level 2, 3, and 4) are able to find out the relations between data in statistical diagrams; 35.6 % of the students (reaching levels of 3 and 4) are able to think critically about the statistical statement; however, only 10.2 % of the students are able to use their knowledge in daily life to give new understanding and explanation to the statistical statement. Therefore, 25.4 % of the students have grasped basic knowledge to carry out critical thinking but have not reached the highest level of critical evaluation of statistical information.

According to the above-mentioned results, three conclusions can be reached. First, more than one-third of the students are not able to carry out critical evaluation on statistical statement. Among all of the students, those with level 0 and level 1 account for 37.5 %. These students are not able to find out the relations between data in statistical diagrams and have not grasped basic statistical knowledge and skills to carry out critical thinking. Second, 25.4 % of the students have the ability to carry out critical evaluation on statistical statement, but their critical evaluation on statistical statement only limits to the question itself. Finally, most of the students (62.5 %) are able to find out the relations between data in statistical diagrams and to carry out some reasoning according to their own understanding, which shows that they have reached basic requirements for statistical learning in senior high school.

12.3.2 *Comparison between Boys and Girls in Levels of Critical Evaluation of Statistical Information*

Table 12.3 shows that there is no significant gender difference at every level. At level 4, data distribution for girls is 3.3 % higher than that for boys. The average score for boys is 14.6, and the average score for girls is 15.2, which is slightly higher than that for boys. Analysis of variance shows that concomitant probability (P)=0.31 > 0.05. There is no significant gender difference statistically.

Table 12.2 The critical thinking levels of the students who received the test

Level	Percentage (%)	Accumulated percentage (%)
4	10.2	10.2
3	25.4	35.6
2	26.9	62.5
1	27.9	90.4
0	9.6	100

Table 12.3 Distribution of critical thinking levels for boys and girls

	Level 0 (%)	Level 1 (%)	Level 2 (%)	Level 3 (%)	Level 4 (%)
Data distribution for boys	10.1	28	27.6	25.9	8.4
Data distribution for girls	9.2	27.7	26.4	25	11.7

Table 12.4 The distribution of critical thinking level for students who participate and don't participate in statistical activities

	Level 0 (%)	Level 1 (%)	Level 2 (%)	Level 3 (%)	Level 4 (%)
Data distribution for students who participate in statistical activities	6.5	25.5	31.5	28	8.4
Data distribution for students who don't participate in statistical activities	12.3	35.9	23	23.1	5.7

Wu's (2004) study on 907 junior high school students' (from 13 to 15 years old in Singapore) understanding ability for statistical diagrams shows that there is no significant gender difference in the interpretation and evaluation of statistical diagrams. Zhang (2007) in her Master's thesis, points out that there is no significant gender difference in the evaluation of statistical diagrams. Tian Zhong's (1999) study also shows that the mathematical thinking ability for boys and girls in junior high school is generally balanced. The author's study confirms the above-mentioned conclusions: there is no significant gender difference in critical evaluation of statistical information.

12.3.3 The Students' Participation in Statistical Activities Can Effectively Enhance Their Critical Thinking Level

Table 12.4 shows that the percentage of students in level 0 and level 1 who participate in statistical activities is 32 %, which is 16.2 % lower than that of the group who don't participate in statistical activities. The percentage of students in level 3 and level 4 who participate in statistical activities is 36.4 %, which is almost 8 % higher than that of the group who don't participate in statistical activities. The students' participation in statistical activities can effectively enhance their critical thinking level.

This conclusion is expectable. Based on different survey questions, groups of students design questionnaires, collect data, analyze data, reach conclusions, write statistical reports, draw statistical posters, and exhibit the results. In this course, students need to use their statistical knowledge to solve practical problems. In order to solve these problems, sometimes they need to better understand the statistical knowledge

they have learned or even need to learn more knowledge by themselves. In this course, students can better understand the effectiveness and limitation of statistics as a tool to process data. Many factors, including questionnaire design, sample selection, the use of statistical methods, and the choice of statistical diagrams, may have direct influence on statistical results. Not all conclusions drawn from statistical survey are reasonable. In order to get scientific conclusions, students need to think about how to design comparatively reasonable survey plan. This process may cultivate students' critical consciousness and enhance their critical thinking capacity.

Besides, members from other groups may raise questions when one group exhibits their survey results. In order to convince the audience, students must think carefully about their survey plan before exhibition, which may also help to enhance their critical thinking capacity.

12.3.4 Analysis of Factors Influencing on Students' Levels of Critical Evaluation of Statistical Information

In order to find out factors that may directly influence on the levels of students' critical thinking, the author makes further analysis on test questions. The author has carefully studied the questions and analyzed the influencing factors. The results are as follows.

According to Table 12.5, the author reaches the following conclusions:

12.3.4.1 The Situation and Background of Statistical Questions May Have Influence on Students' Levels of Critical Thinking

Comparatively speaking, the students did a good job for questions 1 and 2. One-fifth of the students reach level 4. Question 1 is about urbanization, and the students have learned some knowledge about urbanization in their geographic lessons. Question 2 is about computer games, which is an important topic in the education of the students for both parents and teachers. Both situations are close to the everyday life of the students, and they grasp more knowledge about these situations. Therefore, the students are able to think critically about this kind of information.

Question 3 includes the line charts of three kinds of price indexes (real estate price index, producer's price index for manufactured products, and international

Table 12.5 Distribution of levels of students' critical thinking for the questions

Questions	Level 0 (%)	Level 1 (%)	Level 2 (%)	Level 3 (%)	Level 4 (%)
1	1.7	25.8	14.8	35.8	21.8
2	5	16.8	25.8	30.1	22.3
3	9.4	55.9	25.5	5.2	4
4	14.8	30.1	37.6	12.7	4.8

crude oil price index). The students are required to tell if there are relations between the three kinds of price indexes, and they did a quite bad job. Only 9.2 % of the students are able to reach level 3 and level 4, and 65.3 % of them lie in level 0 and level 1. Most students do not understand the statistical situation about how international crude oil price influences producer's price for manufactured products and real estate price; therefore, they are not able to think critically about and give explanation to the conclusion. According to the definition of situation of PISA, the above-mentioned situation falls into "scientific situation" which is comparatively far away from their everyday life. The students are not able to think critically and effectively about this kind of information because they are not familiar about knowledge relevant to this question and situation. The students' answers to this question also reflect the fact that they lack knowledge about social life.

Gal's (2003) study also shows that it is very difficult for the common people to think critically about statistical conclusions when they read online the information and documents (e.g., national economic report, the research report of health information, national import and export information) that they are not familiar about. People's familiarity about the situations of statistical questions may have direct influence on their level of critical thinking.

Data in nature are numbers under certain situation. Situations are the source and base for enhancing understanding, and without specific situation people are not able to explain statistical conclusions. Without relevant situational knowledge, people are only passive receivers of information, and they do not concern about how the data are collected and what calculation and analysis have been carried out. Whether or not the students are able to think critically about statistical information depends on their understanding of situational knowledge. Test results also show that different background of situational knowledge may have influence on students' levels of critical thinking.

12.3.4.2 Different Types of Statistical Diagrams May Have Influence on Students' Levels of Critical Thinking.

Careful analysis of the questionnaire shows that the students' mastery of different statistical diagrams is quite diversified. Different types of statistical diagrams may have influence on students' levels of critical thinking. Generally speaking, the students are good at column diagrams and diagrams with a single variable. They are bad at bar diagrams, data point diagrams with various variables, and scatter diagrams, and sometimes are not able to get useful information from diagrams and give reasonable explanation.

With the development of the economy, it has become one of basic abilities for citizens to understand statistical diagrams and to make decisions based on the diagrams. In statistical teaching, the basic requirement is to teach students to be able to draw and understand statistical diagrams. However, this is far from enough. Teachers should also use examples to teach students basic skills and methods to understand statistical diagrams. In the reading of statistical diagrams, students should not only

get useful information from diagrams but also be able to reach some new conclusions and give their own explanation and hypothesis. In statistical teaching, teachers shall help students to learn by themselves. An important part in statistical teaching is to let students discuss their understanding about some statistical diagrams.

12.3.4.3 Whether or not to Participate in Statistical Survey Activities may Have Influence on Students' Levels of Critical Thinking.

The objective of statistics teaching is to teach students basic methods of statistics and analysis. More importantly, statistics teaching shall let students understand the role and basic thoughts of statistics, know about the difference between thinking of mathematical statistics and thinking of certainty, and notice the randomness of statistical results. Statistical statements can be right or wrong, which depends on the randomness of mathematical statistics. Statistics teaching shall especially emphasize students' participation in statistical process, and let students fully understand thoughts and role of statistics. The most effective method is to let students participate in statistical process.

Through their participation in statistical activities, students have experienced by themselves the process of questionnaire design and data collection. Therefore, they can have better understanding of knowledge about sample selection. For example, in the process of statistical activities, students may realize the randomness of sample. That is to say, when two persons use the same method to treat the same problem, they may get different results because of their different sample selection process. The results have the characteristics of randomness and the conclusion may be wrong. Teachers shall let students realize that although the results may be wrong, statistical deduction is of its significance. It is very necessary for students to think critically about statistical statements. If students have not participated in statistical survey activities by themselves, it is very difficult for them to realize the significance and limitation of statistical deduction. Students' participation in statistical activities is good for the cultivation of their critical thinking ability for statistical statements.

12.4 Teaching Measures and Suggestions

12.4.1 In Statistics Teaching Teachers Shall Teach Students Skills and Methods of Critical Thinking

In order to teach students how to think critically, in statistics teaching teachers shall teach them skills and methods of critical thinking. Gal (2002) designs a table (see Table 12.6) on how to raise critical questions relevant to statistical information.

The purpose of these ten questions is to cultivate students' critical thinking ability. These ten questions, one after another, help students to think critically from the

Table 12.6 How to raise critical questions relevant to statistical information in statistics learning

Raise critical questions relevant to statistical information
Where do these data come from (what is the base for relevant statement)? What type does the study belong to? In the situation, is this type of study reasonable?
Is there any sample used in the study? How is the sample selected? How many people have participated in sample selection? Is the sample big enough? Does the sample include representative groups of people? Is the sample incomplete in certain aspect? In a word, is relevant conclusion drawn from the sample reasonable?
How is the reliability and preciseness of the tools and measures (test, questionnaire, interview) used to get the data of the report?
What is the distribution of the original data (or the data used as the base for statistical deduction)? Is the distribution figure important for the original data?
Is the statistical method used in the report applicable for this type of data? Is it a reasonable model (e.g., using average to summarize conclusion relevant to ordinal numbers)? Does the conclusion using the method of data analysis distort the original meaning of the data?
Is the statistical diagram a proper one? Does the diagram properly describe the trend of data?
How does the probability statement come from? Are there complete and reliable data to verify the likelihood estimation?
As a whole, is the conclusion a reasonable one based on relevant data? For example, is the relationship between data and conclusion close enough? Similar data may lead to quite different conclusions
Is there any other information and method which can help me to understand and evaluate the validity of the conclusion? Is there any information neglected? For example, does the author “habitually forget” the origin of the data for percentage change or sample size?
Is there any other explanation to survey results? For example, is there any intervention or moderator variable with furthering effect which may have influence on the results? Are there any other additional and different meanings which are not mentioned here?

perspectives of statistical knowledge, source of data and conclusions, and to treat all of the conclusions and data in an objective and unbiased way. These ten questions may help students to have an objective understanding of statistical conclusions and may help teachers to cultivate students’ skills of critical thinking in statistics teaching.

12.4.2 In Statistics Teaching, Teachers Shall Consider the Classification of Situations and Diversity of Diagrams When Selecting Sample Questions and Exercises

This study shows that students’ levels of critical thinking are influenced by types of statistical situations and diagrams. In statistics teaching, teachers can choose personal life situation as sample questions, which are more “close” to students’ everyday life. Students are more familiar with knowledge and background of this kind of questions, and it is easier for them to think critically about these questions. Through discussion

among students and summarization by teachers, students can gradually grasp skills and methods of critical thinking and have ability to think critically about statistical statements.

Types of diagrams in sample questions and exercises shall be diversified. Through presentation and discussion of different types of diagrams, teachers shall let students experience three levels of statistical diagrams reading: data reading, reading of the information between data, and reading of the information above data. Students shall be able to read the information between data according to statistical diagrams, to carry out relevant statistical deduction, and to think critically about statistical statements, which shall become an important part of statistics teaching. In this way, students can use their statistical knowledge to think critically about statistical conclusions, to carry out reasonable deduction, and to solve practical problems.

12.4.3 In Statistics Teaching, Teachers Shall Let Students Experience a Complete Statistical Survey Activity.

In order to develop and cultivate students' critical thinking ability, it is very necessary to let students experience a complete statistical survey activity in statistics learning. Through statistical survey activity, students can gradually understand that questionnaire design, sample selection, and the application of analytical methods may all have influence on statistical results. For any statistical survey, we can evaluate the credibility of statistical conclusions only by learning about its survey methodology. Therefore, it is necessary and important to think critically about statistical results.

Besides, by participating in a statistical survey, students can have better understanding of statistical knowledge and methods that they have learned. In the process of solving practical problems, students can gradually change their knowledge into abilities. A precondition for students to be able to think critically about statistical conclusions is to grasp basic statistical knowledge.

What is more, carrying out statistical survey activity in statistics teaching is also a measure to change students' way of learning. When the author organizes students to carry out statistical survey activity, in the process of data analysis, the students have not learned how to use EXCEL to carry out data analysis. The author encouraged students to learn EXCEL by themselves. The students referred to relevant websites to learn by themselves and completed tasks. They provided various types of statistical diagrams. Students from different groups also helped with each other. All of the 22 groups taking part in the survey activity did a good job in data analysis. Therefore, statistical survey activity provides a platform for the development of the students and is good for the cultivation of students' mathematical literacy.

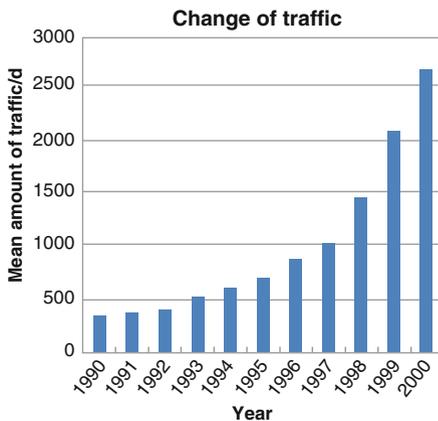
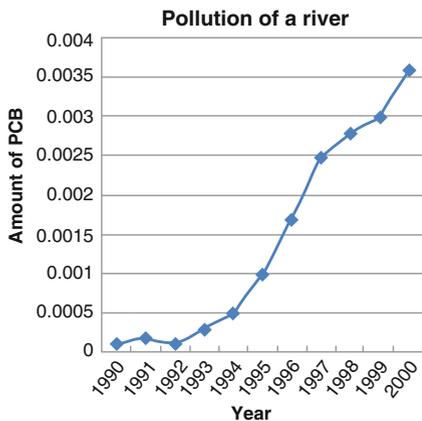
12.5 Limitations of This Study

This study was carried out in the late of June, when grade one students in senior high schools in Shandong Province are preparing for the examination of academic level and terminal examination. After the author carried out questionnaire test, questionnaire coding and analysis, there was no time to carry out interview with students taking part in the test. Therefore, the author cannot make further analysis on the thinking process in the test through in-depth interview, which is a limitation of this study. Coding in this study is based on the results of the questionnaire in written form.

Besides, the sample in this study is grade one students from a key senior high school in Shandong Province. Then how about students from grade two and grade three? Do their critical thinking levels change with different grade? How about students from other regions of China? Is there obvious regional difference? This study cannot answer the above-mentioned questions. There are much more to do in the study of this field.

12.6 Appendix: Test Questions for Senior High School Students' Critical Evaluation of Statistical Information

1. The first diagram shows the pollution of a river in a town in Shandong Province from 1990 to 2000 (PCB refers to the pollution level). The second diagram shows the change of traffic in that town from 1990 to 2000.



According to the two diagrams, someone concludes that the town has urbanized from 1990 to 2000. Do you agree? Please tick the answer which you agree and give your reason.

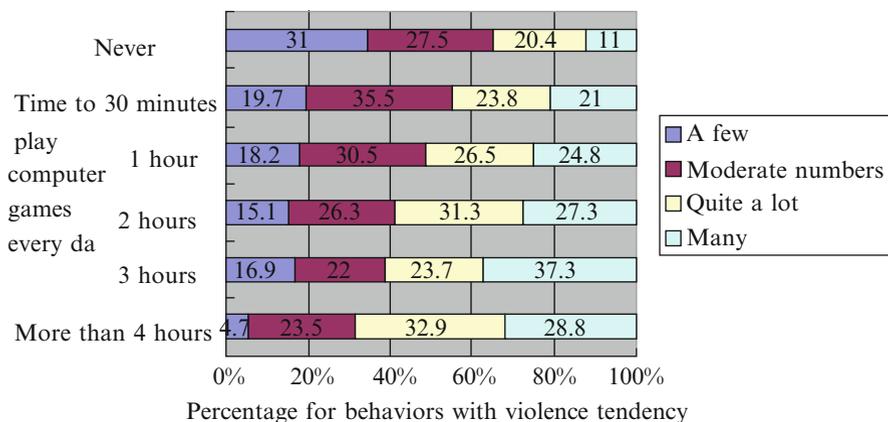
- A. Agree B. Not agree C. Neither A nor B D. I don't know

Your reason:

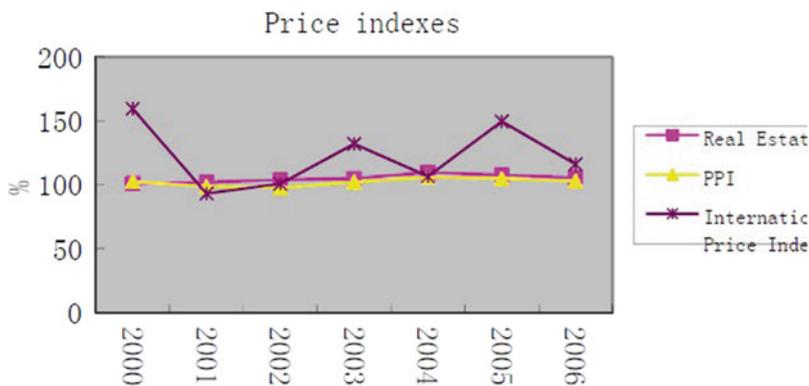
2. The following diagram is the result of a survey for primary school students. The vertical axis shows the time to play computer games every day, and the horizontal axis shows the chance for the occurrence of behaviors with violence tendency (e.g., push or kick classmates, pull other students' hair). (From left to right, there are "a few," "moderate numbers," "quite a lot," and "many.")

Based on the diagram, someone concludes that the reason for the occurrence of violent behaviors among primary school students is that they play computer games for quite a long period of time. Do you agree?

- A. Agree B. Not agree C. Neither A nor B D. I don't know
- Your reason:



3. The following chart shows the change of three kinds of price indexes (the ratio of prices, namely, the ratio of the price of this year to that of last year, so as to describe price change) from 2000 to 2006.



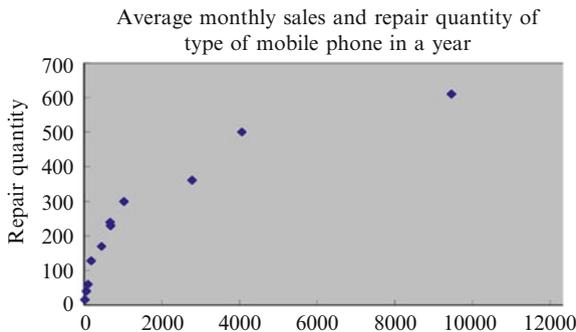
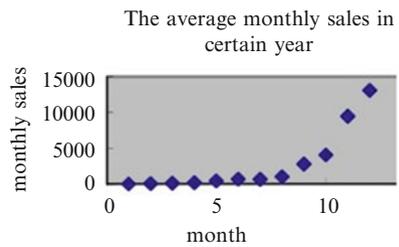
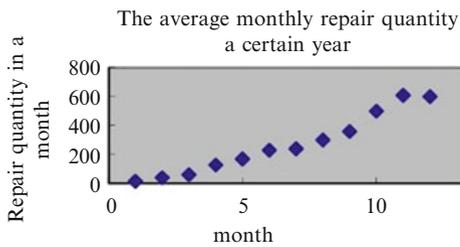
Year	Real estate price index	Producer's price index for manufactured Products (PPI)	International crude oil price index
2000	101.1	102.8	159.5
2001	102.2	98.7	93.4
2002	103.7	97.8	100.8
2003	104.8	102.3	132
2004	109.7	106.1	106.9
2005	107.6	104.9	149.6
2006	105.5	103	116.4

Based on this chart, someone concludes that there is no correlation between the three kinds of price indexes. Do you agree?

A. Agree B. Not agree C. Neither A nor B D. I don't know

Your reason:

4. The following charts show sales and repair quantity of a certain type of mobile phone in a shop.



According to the third chart, someone concludes that the repair quantity of this type of mobile phone is gradually increasing. Therefore, the quality of this type of mobile phone is becoming worse and worse, and consumers should not buy this type of mobile phone. Do you agree?

A. Agree B. Not agree C. Neither A nor BD. I don't know

Your reason:

References

- Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18(5), 382–393. therefore still full of wonder.
- Dan, Z., Jiansheng, L., & Chunli, Z. (2002). The development and training of statistical conception. *Journal of Mathematics Education*, 11(2), 61–63.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities [J]. *International Statistical Review*, 70, 1–51.
- Gal, I. (2003). Expanding conceptions of statistical literacy: An analysis of products from statistics agencies. *Statistical Education Research Journal*, 2, 3–21. Retrieved March 15, 2005, from <http://fehps.une.edu.au/serj>
- Kazuhiro A. (2007). Investigation a hierarchy of students' interpretations of graphs. *International Electronic Journal of Mathematics Education Number3*.
- OECD (2006). *Assessing scientific, reading and mathematical literacy*. A framework for PISA2006. Retrieved from <http://www.pisa.oecd.org/pages/>
- Watson, J. M., & Moritz, J. B. (1997). *Student analysis of variables in a media context*. In Phillips (Ed.), Papers on Statistical Education Presented at ICMI 8 (pp. 129–147). Hawthorn, Australia: Swinburne Press.
- Wu, Y. (2004). *Singapore secondary school students' understanding of statistical graphs*. Paper presented at the 10th International Congress on Mathematics Education (ICME-10), Special Topic Group 11: Research and development in the teaching and learning of probability and statistics. Retrieved from <http://www.icme-organisers.dk/tsg11/Papers/yingkang.doc>
- Zhang, D. (2007). *Survey and research on gender difference of junior high school students' understanding ability of statistical graphs [D]*. Unpublished Master's thesis, Northeast Normal University, Changchun.

Chapter 13

The Conceptual Understanding of Variability in the Data Distributions

Eun-Jeung Ji and Won Kyoung Kim

Recent studies on the concept of variability propose that the idea of spread, or variability should permeate the entire curriculum of statistics, and educators need to modify learning experiences so that students can move comfortably from identifying variability; to describing, representing, and sifting out causes for variability; and finally, to measuring variation (Garfield & Ben-Zvi, 2008). Given this background, the chapter first presents the hypothetical learning trajectory for the concept of variability in brief. This chapter is part of the research carried out for author's doctoral dissertation. The purpose of the original study is to demonstrate a hypothetical learning trajectory for the concept of variability based on local instruction theory as a didactic idea for learning design of the variability concept in the data distributions and sampling situations, theoretically based on the design research.

For the study, we first identified the nature of variability in data distributions and sampling situations. Then we extracted the didactic idea for instruction on the concept of variability from previous studies, the curriculum of the Tasmania in Australia which includes the concept of variability and the materials of AIMS project (Garfield et al., 2010). And then we defined our local instruction theory to teach the concept of variability in the data distributions and sampling situations. Based on our local instruction theory, we developed the hypothetical learning trajectory to be made up of the learning goal that defines the direction, the learning activities, and the hypothetical learning process predicted how the students' thinking and understanding will evolve in the context of the learning activities. According to the hypothetical learning trajectory, we conducted teaching

E.-J. Ji
Daejeon GwanJeo Middle School, Daejeon, South Korea
e-mail: statis333@naver.com (not statis3)

W.K. Kim (✉)
Korea National University of Education, Cheongju, South Korea
e-mail: wonkim@knue.ac.kr

experiments of 14 lessons with ninth-grade students. This chapter shows some examples of the conceptual understanding of variability for 8 ninth-grade students, which is shown in the ways of students describing variability and measuring variation in the teaching experiments.

References

- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, The Netherlands: Springer.
- Garfield, J., delMas, R., Zieffler, A. (2010). *Adapting and implementing innovative material in statistics project*. Retrieved from <http://www.tc.umn.edu/~aims/>

Chapter 14

Students' Misconceptions and Mistakes Related to Measurement in Statistical Investigation and Graphical Representation of Data

Hyung Ju Yun, Eun-Sung Ko, and Yun Joo Yoo

In statistical investigations, the decisions related to the measurement process and the scales of variables can affect the analysis procedures for statistical inference. Students may have difficulty interpreting graphical representations and pursuing statistical inferences due to lack of understanding measurement aspects in data analysis. Misconceptions and misunderstandings related to measurement could arise in various areas in statistical investigations. To measure, proper operational definitions for variables should be given. The consistency of measurement should be maintained. The level of measurement matters. Measurement error should be controlled. For histogram representations, the decision on the number of classes is required and the continuous variables are transformed into categorical variables resulting in some information lost. In this study, we developed a questionnaire to investigate the misconceptions related to measurement in above areas and analysed 213 middle school students' responses to identify how often these misconceptions are observed.

Students were asked to judge about the data within several example situations such as investigations on the distance between school and students' house, foot length, eyesight, amount of water drunk, math exams and broad jump record. Related to operational definition, 21.6 % students thought measuring the 'time to get to school' is enough to investigate the distance between school and students' house instead of obtaining 'house address' (52.6 %). Only 8.0 % of students found

H.J. Yun

Graduate School of Seoul National University, Gwanak-gu, South Korea
e-mail: pearlia@snu.ac.kr

E.-S. Ko

Jeonju National University of Education, Jeonju, South Korea
e-mail: kes7402@jnue.kr

Y.J. Yoo (✉)

Seoul National University, Gwanak-gu, South Korea
e-mail: yjoo@snu.ac.kr

it problematic using shoe size to learn about the foot length. For the consistency of measurement, 96.7 % students did not find the problem in mixing the values from two different measuring tools (Snellen chart testing and power of lens) resulting different units mixed in the data of eyesight and 95.8 % did not find the problem in using different sized cups to measure the amount of drunk. For the level of measurement, 60.6 % students could choose proper graph representation between histogram and bar chart for various continuous and categorical variables but for shoe size variable 63.4 % thought histogram is more appropriate even though the number of different values is only six. Only 12.7 % of students were hesitant to use one-time measurement to decide on the ability for broad jump recognizing the possibility of measurement error. For the number of classes in the histogram, students showed fixation on the number of classes to be around 6–8 and the look of the graph to be familiar to the ones shown in school texts (65.3 %) even though the other histogram with many classes (29.6 %) provides accurate information for the purpose of the investigation. 37.6 % students thought the distribution of the values in each interval of a histogram is uniform and tried interpolation to obtain a conclusion on the middle value within an interval.

In conclusion, we found various misconceptions and misunderstanding occurring in many students even though these are related to basic characteristics of measurement in statistical investigation. We suggest students be taught about statistics and graphical representation with more stress on the context of measurement process and scales of measured variables.

Chapter 15

Overall Understanding of the Middle School Mathematics Course in Teaching Statistics

Main Line

Guan Jian, Bai Xue, Wu Peng, and Li Qiusheng

Compared with statistical teaching in foreign countries, it is relatively late in China. At present, the number of the research on the development of the case teaching statistics is very little in China. On the base of the analysis of Chinese obligation education stage and high school math curriculum standard and by comparing domestic and international mathematics curriculum standards in statistics, we find that there are no concrete cases in statistical teaching in China, although the curriculum standard explicitly points out the case of statistics is so important. From the “overall understanding of the middle school mathematics course teaching statistics main line” point of view, this chapter describes our understanding of the teaching statistics main line in middle school mathematics course with the guidance of statistics thinking.

G. Jian (✉)

Beijing Yuying School, Haidian, China
e-mail: janejoyce50@126.com

B. Xue

Haidian the teacher training school, Haidian, China
e-mail: hdjx_bx@sina.com

W. Peng

Beijing JingShan School, Dongcheng, China
e-mail: kenanwupeng@sina.com

L. Qiusheng

Beijing Attached middle school of Chinese People’s University, Haidian, China
e-mail: liqs@rdfz.cn

First, we defined statistical thinking from the angle of philosophy and statistics. Then this chapter provides some typical statistical cases which are of practical value. After that it introduces statistical teaching case of elementary school, junior high school, and senior high school. Finally, we analyze the cases of the statistics from the angle of philosophy and statistics.

This research provides a more feasible pattern for China's statistics teaching. On the base of analyzing a few classic cases, we formed the view of overall understanding of the middle school mathematics course teaching statistics main line, which reflects statistical thinking. This chapter suggests middle school mathematics teachers should grasp middle school mathematics course teaching content of statistics as a whole. They should help students to think in statistical way and to use it in real life.

Statistics → statistical thinking → curriculum standard → the overall assurance

References

- Chen, X. (2002). *The brief mathematical statistics*. Changsha, China: Hunan Education Press.
- Darrell, H., & Liao, Y. (Trans.). (2002). *How to lie with statistics* (p. 5). Shanghai, China: Shanghai University of Finance & Economics.
- Gudmund, R. I., & Gergen, M. (1997). *Statistics the conceptual approach* (X. Wu, Trans.). CHEP. New York: Springer.

Chapter 16

Growth on Fourth-Grade Students' Mathematical Understanding of Average

Jun Wu

Many articles had discussed students' various difficulties of understanding of average from different aspects. However, as a complex psychological phenomenon, understanding is interior and indirectly observed. It is not easy to reveal the nature and development of it. How do we present the development process of understanding?

We used Pirie–Kieren dynamical model (Pirie & Kieren, 1994) to analyze students' growth of mathematical understanding of average. The instrument is Hats Averaging Problem (Cai, 2000) (see Fig. 16.1), and we interviewed three fourth-grade students A, B, and C in a primary school in a city. It could be presented with a sketch map in which a curve described student C's development of understanding of average (see Fig. 16.2). Comparatively speaking, student A and B had extra experience of folding back from property-noticing level to image-making level and turned back again.

J. Wu (✉)

School of Mathematics and Information Science, Qujing Normal University, Qujing, China
e-mail: ynwujun@126.com

Week 1	
Week 2	
Week 3	
Week 4	?

Fig. 16.1 Hats averaging problem

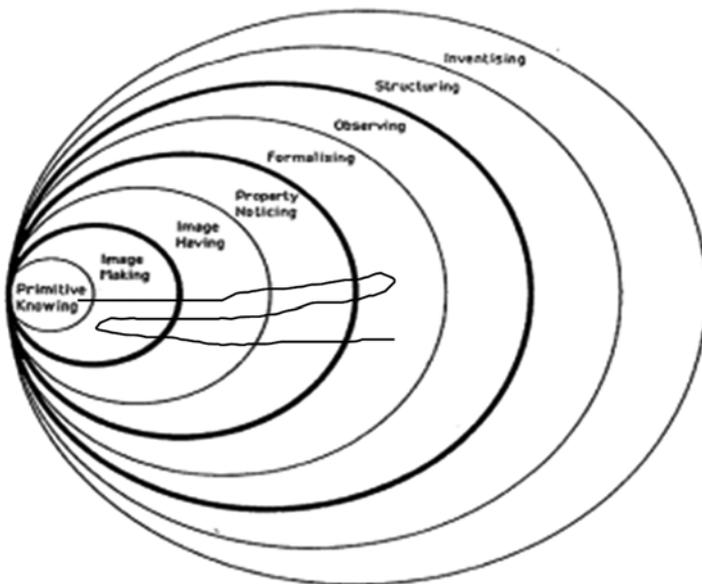


Fig. 16.2 Pirie-Kieren model

References

Cai, J. (2000). Understanding and representing the arithmetic averaging algorithm: An analysis and comparison of US and Chinese students' responses. *International Journal of Mathematical Education in Science and Technology*, 31(6), 839-855.

Pirie, S. E. B., & Kieren, T. E. (1994). Growth in mathematical understanding: How can we characterize it and how can we represent it? *Educational Studies in Mathematics*, 26(2-3), 165-190.

Chapter 17

Exploring Dot Plot in the Perspective of Embodied Cognition

Verônica Yumi Kataoka, Irene Cazorla, Eurivalda Santana,
and Claudia Borim da Silva

The effective inclusion of Probability and Statistics teaching in Brazilian schools faces great challenges: teachers' training and a lack of didactic materials and software among others (Cazorla, 2006). In this teaching context, it is common to observe, among the students and even in textbooks, conceptual errors and not explore some graphs, for example, the dot plot and the box plot. Reflecting on these questions, in 2008, a group of Statistics educators started to develop the Virtual Environment for Supporting Statistics Literacy in Basic Education (AVALE-EB), which is a free virtual learning environment that currently offers ten didactic sequences (DS) to help Probability and Statistics teachers at school, both in a pencil and paper environment and in a virtual environment.

One of the graphs that is explored in the most DS is the dot plot since it is considered a graph with a high didactic potential to help students to understand mainly the concept of variability, but also it possible the immediate visualization of the mode, maximum and minimum values, total range. The objective of this chapter is to present theoretical aspects of the human and paper dot plot exploration in the perspective of embodied cognition. According to this theory, cognitive processes not only depend on brain activity but are also linked with different parts of the human body (Lakoff & Johnson, 2002).

V.Y. Kataoka (✉) • I. Cazorla • E. Santana
Universidade Estadual de Santa Cruz, Ilhéus, Brazil
e-mail: veronicayumi@terra.com.br; icazorla@uol.com.br; eurivalda@hotmail.com

C.B. da Silva
Universidade São Judas Tadeu, São Paulo, Brasil
e-mail: dasilvm@uol.com.br

In order to construct the human dot plot, the students are asked to line up at the front of the classroom or in the school playground, according with the variable of interest, for example, to measure the variable height the students are asked to line up according with their height, and it is natural for the students with the same height to form perpendicular lines (piling dots). The students have the chance to witness the tendency of data using their own bodies. After that the students draw the dot plot on paper. It has been observed when working with students that the human dot plot is a graphic representation that seems to be representative to them and how it becomes less abstract.

References

- Cazorla, I. M. (2006). Teaching statistics in Brazil. In: A. Rossman, & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. International Association for Statistics Education, Salvador (Brazil). CD-ROM.
- Lakoff, G., & Johnson, M. (2002). *Metáforas da vida cotidiana* (M. S. Zanotto & V. Maluf, Trans.). São Paulo: Mercado das Letras.

Part III
Learning Statistics at the Tertiary Level

Chapter 18

Students' Difficulties in Understanding of Confidence Intervals

Ana Henriques

18.1 Introduction

The teaching and learning of statistics has acquired a great importance due to its recognized role in the general education of citizens (NCTM, 2000). This increasing attention has resulted in the development of research related to students' difficulties with statistical concepts which should guide the construction of teaching situations to overcome the cognitive obstacles identified (Garfield & Gal, 1999; Shaughnessy, 2007).

The concept of confidence interval (CI) is central in introductory courses in statistics, both in secondary education and at university. However, as a teacher of statistics, there is plenty of evidence from my own experience that the proper understanding of confidence intervals seems to be unusually difficult, a view which is confirmed by several research studies (Callaert, 2007; Canal & Gutiérrez, 2010; Olivo, 2008; Shaughnessy, 2007).

This study aims to describe and analyse the students' difficulties in understanding of confidence intervals and to determine whether they are specific to the learning environment of students or are common to those mentioned in other studies. The study also aims to develop a more elaborated view concerning the presence of the diverse elements of meaning in students' responses which reveal the concept comprehension (Godino & Batanero, 1994). In this sense, this study extends and deepens the previous works about students' understanding of confidence interval and thus contributes to the dissemination of research results related to the conceptual difficulties of students, which are not yet sufficiently known to teachers and should be taken into account to improve the teaching and learning of this topic.

A. Henriques (✉)

Instituto de Educação, Universidade de Lisboa, Lisboa, Portugal

e-mail: achenriques@ie.ul.pt

18.2 Theoretical Framework

The confidence interval is one of the general procedures of statistical inference that can be applied to various fields of problems (Callaert, 2007; Fidler & Cumming, 2005; Olivo, 2008). The most commonly referred are: Estimation of unknown parameters of a population, comparison of distributions, hypotheses testing of population parameters, determination of the appropriate sample size to make an inference and determination of the limits of tolerance. Each of these fields is very broad and varied and includes the estimation of averages, proportions, variances and regression and correlation parameters. Depending on the parameter, we also find many different conditions and assumptions (cases of known or unknown variance, equal or different variances, large or small samples and normal or non-normal populations). These assumptions determine the sampling distribution to be used in the construction of the confidence interval, and therefore there is a wealth of probability distributions that students should learn and relate to the conditions of the problem (e.g. normal, t -student, chi-square and F). Procedures, such as computing probabilities and critical values, linear or functional transformations, are also necessary. The confidence interval is thus considered a complex concept and difficult to understand since it involves various problems and situations from which the concept arises, different representations of the concept, definitions and properties, procedures and strategies to solve problems and arguments that the student must know in advance to validate propositions.

Godino and Batanero (1994) define five interrelated components in the meaning of a mathematical object that should be specifically dealt with in organizing instruction or in assessing students' learning: *Extensive elements*—fields of problems from which the concept arises, the context from where it is induced and to where the CI is applied; *Ostensive elements*—representations of the concept used in the mathematical activity (graphics, tables, notations, expressions, terms); *Actuating elements*—procedures and algorithms to solve problems or to compute its values; *Intensive elements*—definitions of the concept, its proprieties and relationships to other concepts and *Validating elements*—type of arguments and proofs used to validate solutions and propositions.

Confidence intervals are not always properly interpreted and are prone to misconceptions. Although students might be able to perform all necessary manipulations and formal calculations to construct confidence intervals, it has been shown that many of them hold deep misconceptions that have a direct impact on learning inferential statistics because of the interconnection of the concepts and methods and the relevance of their understanding for an appropriate interpretation of inferential results and conclusions. Thus, the difficulties experienced by students in understanding the confidence intervals were the focus of previous research. A very common difficulty is to interpret confidence intervals as a statistical descriptive object rather than a statistical inferential one providing information about the value of a population parameter (Canal & Gutiérrez, 2010; Fidler & Cumming, 2005). In other studies (Behar, 2001; Fidler, 2005) it has also been reported that students are not clear on the direct or inverse relations between the interval width, the sample

size and the confidence level. Olivo and Batanero (2007) identified other difficulties involving errors in determining the critical values or the selection of the adequate sampling distribution. These studies also made clear that much more research is still needed to shed light on the sources of those difficulties, clarifying the fundamental components in the meaning of this specific concept since students might have difficulties in all of them.

18.3 Method

In order to assess students' difficulties in understanding of confidence intervals, a written test was applied to 33 second-year students (4 female and 29 male) of the Naval Academy who attended an introductory statistics course taught by the author through traditional learning situations. Students answered the test individually in one of the last classes of the semester, after the topic had been addressed. The test consisted of five multiple choice items, in which students were asked to select the single correct answer and four problems for which students were encouraged to give detailed explanation to their answers. The questions that comprised this test were obtained from Olivo (2008). Due to restrictions in length, in this chapter I mainly present: (1) the quantitative results for the five multiple choice items focus on the conceptual nature of the CI and the relations between their width and the level of confidence, the sample size and the variability of the population; and (2) a qualitative analysis of students' difficulties in solving four problems concerning the procedural knowledge in the construction of CI. The students' answers for each problem were classified as correct, partially correct or incorrect based on previously developed criteria of Olivo (2008). That classification fits the needs raised from the nature of the responses in this study and allows a comparative analysis with the results obtained by the author. I also use the theoretical model of Godino and Batanero (1994) to describe the elements of meaning present in the students' responses.

18.4 Results and Discussion

Table 18.1 shows a summary of the answers to the multiple choice items in the test, the proportion of correct answers and the description of their content. It is observed that all items were solved correctly by more than half of students and that the average percentage of correct answers was approximately 65 %.

The students revealed some difficulties in the definition of confidence interval. Although 70 % of students have answered item 1 correctly, we have identified errors in understanding the inclusion of the population parameter in the interval in 24 % of the answers and in 6 % of the answers, students do not understand its inconsistency considering the sample.

A number of facts stand out from the analysis of the answers to items 2–5, which focus on relations between interval width and confidence level, sample size and

Table 18.1 Students' answers to multiple choice items

Frequency item	a	b	c	d	Correct answers (%)	Item content
1	2	5	23	3	70	Confidence interval definition
2	0	26	3	4	79	Relation of interval width and sample size
3	0	16	17	0	52	Relation of interval width and confidence level
4	7	18	3	5	55	Relation of interval width and confidence level
5	1	22	5	5	67	Relation of interval width and population variability

Table 18.2 Students' answers to problems

Problem	C	PC	I (%)	Field of problem
The average of 100 classifications in a mathematics test is 75. Assuming $\sigma=7$, find a 95 % CI on μ	97	0	3	Construction of CI on μ , large σ
We have obtained the following data for the daily emission of sulphur oxides, for a sample of 100, $x=18$ and $s^2=36$. Find a 95 % CI for the real daily average emission of sulphur oxides	97	0	3	Construction of CI on μ , large sample, unknown σ
Find a 95 % CI for the mean value of a normal population with unknown standard deviation if a random sample of 10 gives $x=25$ and $s=6$	61	0	39	Construction of CI on μ , small sample, unknown σ
Let σ^2 be the variance of disruptive tension distribution. The sample variance obtained for $n=16$ is $s^2=13,700$. Find a 95 % CI on σ	49	33	18	Construction of CI on σ

population variability. About 80 % of the students know the correct relation between precision and sample size. However, for 50 % of the students it is not clear how the confidence level of the interval affects its width. In item 3, the answers are divided between the correct answer and the answer that is contrary to that same one, thus revealing that students recognize the influence of the confidence level in the interval width but they find it difficult to identify the trend of that relation (direct or inverse). In item 4, 21 % of the students indicate a relation in the same direction between the width of the interval and the sample size, confirming the difficulties which have occurred in the previous item. A lower percentage of students (9 %) revealed difficulties in understanding the effect of the population variability in the interval width. This result is consistent with the answers in item 5 in which the percentage of correct answers is 67 %. The difficulties in this item are divided between considering that the variability increase decreases the interval width or makes it the same. Therefore, we can say that students revealed many difficulties in questions concerning factors which influence the interval width.

Table 18.2 presents a summary of the students' answers to the problems in the test, including their written statements, the percentage of correct answers (C), partially correct answers (PC), and incorrect answers (I), as well as the description of their content. In order to complete this table, I am also analysing the students' answers to the problems, concerning the difficulties which were observed and also the presence of several elements of meaning.

Items 1 and 2 stand out due to the near absence of incorrect answers which, according to Muñiz (1994), indicates how easily students deal with their resolution. In these items, all students but one symbolically represent the problem data, which refer to various concepts—sample size, sample or population standard deviation, degrees of freedom and confidence level (ostensive and intensive elements were used), and they have correctly identified the population parameter to be estimated—mean value, in order to justify the choice of the sampling distribution to be used (intensive and validating). In item 1, these students used the normal distribution correctly. In item 2, the correct answers, were divided between the 51 % who used the normal distribution, justifying that sample size is greater than 30, the 30 % who used the normal distribution without justifying (this seems to be implied as these students, in item 3, do not reveal misconceptions regarding the choice of distribution when σ is unknown) and the 19 % who have focused on the σ unknown, and they do not take into account the sample size and select the student-t distribution (intensive and validating). However, that does not condition the solution because the table that was available for this distribution does not include the 99 degrees of freedom and students make an approximation to 100, thus obtaining the critical value corresponding to the normal distribution and ending up getting the same interval (intensive and ostensive).

Therefore, the students correctly determine the critical values of the normal distribution or the t -student corresponding to the 95 % confidence level by subtracting this value to the total probability (1) and dividing the remainder 0.05 into two equal parts (intensive and actuative). Some students feel the need to represent the distribution graphically to help them with this calculus (ostensive). Even though not everybody remembers to write the CI expression, they replace the problem data in the expression and, by using deductive reasoning, they perform several algebraic operations with inequations in order to find the confidence interval (actuative and intensive). In the Fig. 18.1 below, we can find examples of the situations mentioned which were considered correct answers:

The table only provided data for degrees of freedom equal to 100 and not to 99, but by approximation values corresponding to 100 freedom degrees are used.

$\Phi(x) = 0.975$
 $x = 1.96$
 $\theta = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim N(0,1)$
 $N = 100$
 $\bar{Y} = 18$
 $S^2 = 36$
 $S = 6$
 $\alpha = 5\%$
 $\frac{\alpha}{2} = 2.5\%$
 $P[-1.96 \leq \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \leq 1.96] = 0.95$
 $P[-1.96 \leq \frac{18 - \mu}{\frac{6}{\sqrt{100}}} \leq 1.96] = 0.95$
 $P[-1.96 \cdot \frac{6}{10} \leq 18 - \mu \leq 1.96 \cdot \frac{6}{10}] = 0.95$
 $P[-1.176 \leq 18 - \mu \leq 1.176] = 0.95$
 $P[16.827 \leq \mu \leq 17.173] = 0.95$

$S^2 = 36 \Rightarrow S = 6$
 $n = 100 > 30$
 usamos una distribución normal.
 $P[-1.96 \leq \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \leq 1.96] = 0.95$
 $P[-1.96 \cdot \frac{6}{10} \leq 18 - \mu \leq 1.96 \cdot \frac{6}{10}] = 0.95$
 $P[-1.176 \leq 18 - \mu \leq 1.176] = 0.95$
 $P[16.827 \leq \mu \leq 17.173] = 0.95$

Fig. 18.1 Examples of correct answers to item 2 (using an approximation to $N(0,1)$ and justifying the choice of $N(0,1)$)

Only one student incorrectly answered these two items because he made an error in reading the table of the normal distribution (intensive and ostensive) and possibly he transfers that value from one item to the other, without using the table again. When looking for the critical value corresponding to a probability of 0.025, the student used the value 0.25 by distraction and then he confused the critical value with the respective probability (he looked for $P[X=0.25]$ instead of $\Phi^{-1}(0.025)$). Some notational inaccuracies (ostensive) were also observed and he didn't finish the problem as shown in Fig. 18.2:

In item 3, it was expected that students understood the influence of the sample size in the selection of the sampling distribution when the σ is unknown. Only 61 % of students answer correctly, constructing a confidence interval by selecting the t -student distribution. These students evoke the population parameter which they want to estimate (mean value) and the fact that σ is unknown and sample size less than 30 in order to justify the choice of sampling distribution to be used (intensive and validating). In none of the answers do we see errors in the determination of the degrees of freedom. The difficulties identified in these answers are associated with the notation (ostensive), although this does not affect the solution. The incorrect answers were originated in two different situations. Seven students (21 %) showed difficulty in obtaining the critical values from the table, by considering 5 % for each of the tails of the t -student distribution rather than 2.5 % (Fig. 18.3):

This result is surprising due to the fact that it has not occurred in previous items. In addition, the difficulty is not related to the concept of CI, but it originates from the lack of understanding of what the critical value is (intensive). The other incorrect

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$n=100 \quad \bar{X}=75 \quad \sigma=7$$

$$IC=0.95 \quad 1-\frac{\alpha}{2}=0.015$$

$$P(0.015)=0.98$$

$$P\left[-0.68 \leq \frac{75 - \mu}{\frac{7}{10}} \leq 0.68\right] = 0.95 \Leftrightarrow$$

Fig. 18.2 Example of incorrect answer to item 1

$$Q_2. \quad \alpha = 1 - 0.95 = 0.05$$

$$m = 10$$

$$\bar{X} = 25$$

$$n = 6$$

$$\sigma = \text{Desconhecido}$$

$$\mu = ?$$

$$z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{m}}} \sim t(9)$$

$$P[-1.833 \leq \frac{25 - \mu}{\frac{6}{\sqrt{10}}} \leq 1.833] = 0.95 \Leftrightarrow$$

$$\Leftrightarrow P[-3.478 \leq 25 - \mu \leq 3.478] = 0.95 \Leftrightarrow$$

$$\Leftrightarrow P[-28.478 \leq -\mu \leq -21.522] = 0.95 \Leftrightarrow P[21.522 \leq \mu \leq 28.478] = 0.95$$

Fig. 18.3 Example of incorrect answer to item 3

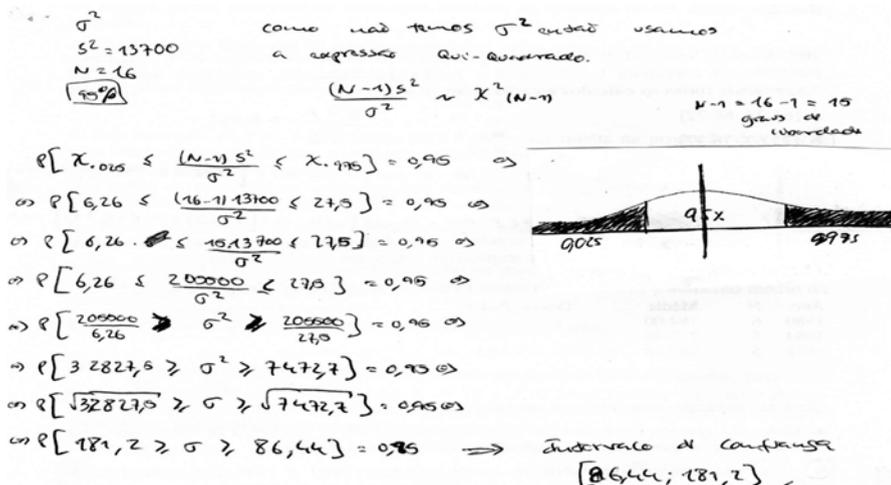


Fig. 18.4 Example of a correct answer to item 4

answers (18 %) do not originate from errors made in applying the procedure of constructing intervals but, possibly, either from the lack of careful reading of the written statement or from the confusion with item 4 that followed since all of these students constructed a confidence interval for the variance. Thus, they have not answered as required.

Item 4 was the most difficult one for the students since only 49 % of them answered correctly. In these answers, students identify the population parameter to be estimated (standard deviation), and they justify their choice of the sampling distribution for the variance (intensive and validating). Most of these students represent graphically the χ^2 distribution (ostensive). However, some students show a conflict related to their representation (ostensive) because they draw a symmetric graph, as they are probably confusing the normal distribution, as shown in the example of Fig. 18.4. However, this does not affect the solution since they use the Chi-square table to find the critical values of χ^2 corresponding to the confidence level of 95 % (ostensive and actuative).

For the answers considered partially correct, several difficulties were identified. Seven students did not finish the problem since they seem to have forgotten the last step and to provide an answer for the problem indicating a confidence interval for the variance instead of the standard deviation. Four students still made procedural errors in the resolution of the inequations (actuative) due to the fact that the parameter to be estimated is in the denominator in the expression of the random variable, although the entire previous process of the interval construction was carried out with no mistakes (Fig. 18.5):

For the answers considered incorrect (18 %), one of the students selected the normal distribution (intensive), but he did not perform any further procedure, ultimately not answering the problem. One other student mentions the Chi-square distribution and other concepts—sample size, variance and degrees of freedom (intensive)—but

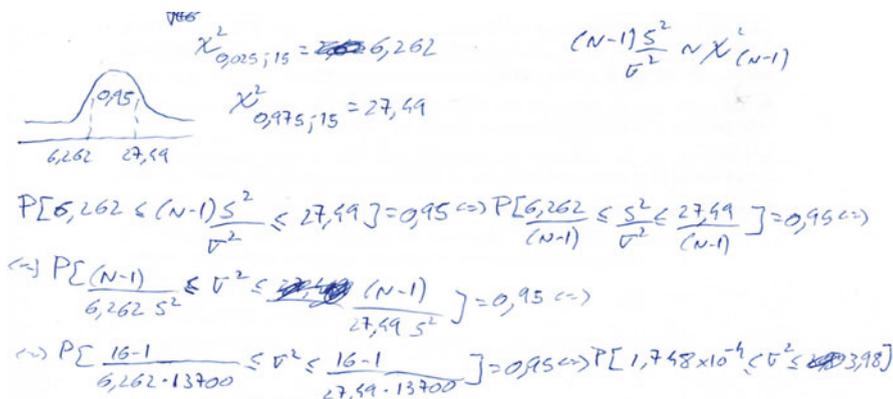


Fig. 18.5 Example of partially correct answer to item 4

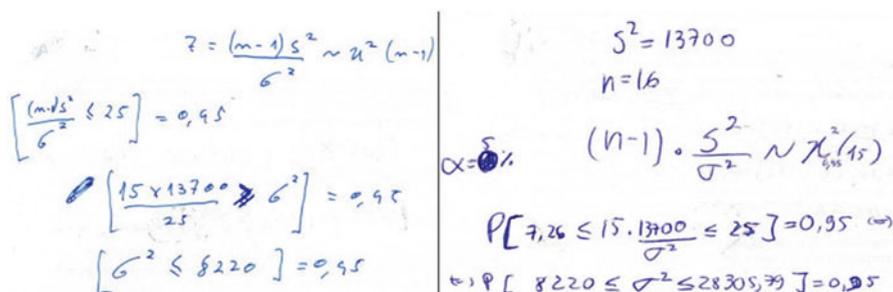


Fig. 18.6 Examples of incorrect answers to item 4

he writes one formula incorrectly (ostensive), thus displaying notational conflicts, as he equals the interval and the probability and he seems to recall only one limit. Therefore, he is not answering to as required. One other difficulty presented by some students (4) was the calculus of the critical value of the distribution related with the level of confidence of the CI, as already seen in the previous item. Students use the table correctly to find the critical values of χ^2 , but they make a procedural error by replacing them by the positions corresponding to the confidence level of 90 % (instead of 95 % as required), and not considering the bilateral interval (ostensive and active). The following excerpts are examples of that and also show other difficulties already described (Fig. 18.6):

In short, the main difficulties encountered regarding the understanding of elements of meaning that were deduced from the analysis of answers to problems are:

Extensive elements. We can observe difficulties in the selection of the appropriate sampling distribution to construct a confidence interval for the mean value when σ is unknown which will influence the confusion between problem fields and the corresponding application of incorrect procedures for calculating intervals.

Intensive elements. Some students had difficulty in obtaining the critical values from the table taking one tail area which was double of the correct one, resulting in a CI of less confidence than the required. In addition, they do not always discriminate the cases of known and unknown population standard deviation by selecting the appropriate sampling distribution to construct the CI.

Actuating elements. Students both understand and find it easy to apply the algorithm to find the CI. The resolution of inequalities in order to calculate the CI when it is necessary to perform inversions to find the σ^2 seems to be the hardest for students. The careless reading of the written statement or their distraction has also caused many students not to finish problem 4, presenting a CI for σ^2 instead of σ .

Ostensive and Validating elements. Concerning the symbolic language, difficulties consist of: they do not write the indexes $1-\alpha/2$ and $\alpha/2$ in the critical values of distributions; they forget to use the P (representing probability) and they often equal the interval to 95 % or they use the P without indication of the variable (e.g. $P[0.025]=0.68$); they do not always present the CI in the form of interval; and they make mistakes in the expression of the CI to the σ^2 presenting only one side. One student also confused the critical value with the respective probability when reading the table. There was also some confusion between the graphs of normal and t -student distributions (symmetric) and the χ^2 distribution, which students presented with the same symmetry allowing for these to take negative values. There is a tendency for students not to explicit the arguments which they have used, despite these being suitable in most cases.

18.5 Conclusions

The results suggest that, in general, students understand the concept of confidence interval. Students recognize both the problem fields and the properties of the concepts used in the construction of the CI, and they are able to perform the algorithm correctly. However, that understanding is only apparent since they made many errors and revealed some difficulties in relation to confidence intervals (conceptual, procedural and interpretative) that reinforce the conclusions of previous investigations. The answers from multiple choice items show that the definition of confidence interval proved to be a difficult issue for students. As in Fidler and Cumming (2005) and Behar (2001), students tend to interpret confidence intervals as a statistical descriptive object. The most frequent errors occurred when students had to interpret the relations between width and other elements associated with the concept (sample size, confidence level and population variability). There were also difficulties in the selection of the adequate sampling distribution, in particular related with the construction of CI on variance and in the determination of critical values, which may have to do with the lack of students' understanding of probability distributions as referred in Olivo and Batanero (2007).

The analysis of the elements of meaning in the problems confirmed the difficulties of some students in relation to the CI which can be explained by the existence of a wide variety of semiotic conflicts to be taken into account when organizing teaching. Consequently, these results reflect the need for dedicating more time to the topic and for changing teaching approaches at university.

Although students in this study perform better, the results and the observed difficulties are consistent with those found in Fidler (2005) or in Olivo and Batanero (2007) with students from the same level of education. Thus, the results seem to indicate that students' difficulties with the confidence interval concept are not specific to the education systems and highlight the need for further research, with large groups of participants from different backgrounds in order to present empirical evidence about misconceptions and to develop pedagogical tools to confront students with their difficulties and to find possible means to help students to overcome them.

Acknowledgments This study was supported by national funds through FCT – Fundação para a Ciência e Tecnologia, Project Developing statistical literacy: student learning and teacher training (contract PTDC/CPE-CED/117933/2010).

References

- Behar, R. (2001). *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística*. Tesis doctoral, Universidad Politécnica de Cataluña, Barcelona, Spain.
- Callaert, H. (2007). Understanding confidence intervals. In *Proceedings of the 5th Congress of European Society for Research in Mathematics Education* (pp. 692–701). Larnaca, Cyprus.
- Canal, G. Y., & Gutiérrez, R. B. (2010). The confidence intervals: A difficult matter, even for experts. In *Proceedings of the 8th International Conference on Teaching Statistics*. Ljubljana, Slovenia. Retrieved from www.stat.auckland.ac.nz/~iase/publications.php
- Fidler, F. (2005). *From statistical significance to effect estimation*. Doctoral Thesis, University of Melbourne, Melbourne, Victoria, Australia.
- Fidler, F., & Cumming, G. (2005). *Teaching confidence intervals: Problems and potential solutions*. Retrieved from www.stat.auckland.ac.nz/~iase/publications.php
- Garfield, J. B., & Gal, I. (1999). Teaching and assessing statistical reasoning. In L. Stiff (Ed.), *Mathematical reasoning in grades K-12* (pp. 207–219). Reston, VA: NCTM.
- Godino, J. D., & Batanero, C. (1994). Significado personal e institucional de los objetos matemáticos. *Recherches en Didactiques des Mathématiques*, 14(3), 325–355.
- Muñiz, J. (1994). *Teoría clásica de los tests*. Madrid, Spain: Ediciones Pirámide.
- National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Olivo, E. (2008). *Significado de los intervalos de confianza para los estudiantes de ingeniería en México*. Tesis doctoral, Universidad de Granada, Granada, Spain.
- Olivo, E., & Batanero, C. (2007). Un estudio exploratorio de dificultades de comprensión del intervalo de confianza. *Unión*, 12, 37–51.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Greenwich, CT: Information Age/NCTM.

Chapter 19

A Modeling and Simulation Approach to Informal Inference: Successes and Challenges

Jennifer Noll, Mulugeta Gebresenbet, and Erin Demorest Glover

19.1 Introduction

Nolan and Lang (2010) argue that computational literacy is now “fundamental to statistical practice... vital to all facets of a statistician’s work... yet it occupies an astonishingly small proportion of the statistics curricula” (p. 98). Cobb (2007) argues that computer technology offers statistics educators an opportunity to place more emphasis on the key concepts of inference (i.e., chance models and determining statistical unusualness) and less emphasis on procedures (i.e., formulaic hypothesis tests like z and t -tests). He states, “we may be living in the early twenty-first century, but our curriculum is still preparing students for applied work typical of the first half of the twentieth century” (p. 7). Based on the recommendations of these statisticians, educators are developing new approaches to the teaching of statistics that align with recent trends in the way statisticians work with data. In the past 2 years, new curricula have emerged that focuses on randomization techniques and computer simulations (e.g., Change Agents for Teaching and Learning Statistics (CATALST), Garfield, delMas, & Zieffler, 2012; Morgan, 2011). The arguments made by statisticians and educators to update and develop new technology-driven curricula are only

J. Noll (✉) • E.D. Glover
Portland State University, Portland, OR, USA
e-mail: noll@pdx.edu; eglover@pdx.edu

M. Gebresenbet
Addis Abeba University, Addis Ababa, Ethiopia
e-mail: mwolde99@yahoo.com

one component of improving introductory statistics. If we are to change the “culture” of the statistics classroom, aligning it more closely to the way statisticians *do* statistics, a better understanding of how technology impacts students’ development around the logic of inference is needed. This chapter presents an empirical investigation of introductory statistics students’ developing informal conceptions of inferential statistics as they engage in modeling and simulation using TinkerPlots™ (TP).

19.2 Background

Working with samples of data and using samples to make inferences about unknown populations are key components of statistical investigations. Research on students’ informal inferential reasoning suggests that students have many difficulties in understanding and using statistical inference, including building a schema of many interrelated ideas such as representativeness, sampling variability, and distribution (Saldanha & Thompson, 2003).

Rubin, Bruce, and Tenney’s (1991) research revealed a tension among students between being focused on sampling representativeness and sampling variability.

Many statistics educators now advocate teaching inference from an empirical perspective through simulation, which they argue helps students better understand how statistical decisions are made (Chance, Ben-Zvi, Garfield, & Medina, 2007; Chance, delMas, & Garfield, 2004; Cobb, 2007; Garfield & Ben-Zvi, 2009). From an empirical approach, the study of sampling variability typically focuses on taking repeated samples from a population, creating a distribution of sample statistics from those repeated samples (such as sample means or sample proportions), and comparing the observed sample statistic from the research to the empirical sampling distribution. This comparison allows students to see how the observed sample statistic compares to the distribution of sample statistics created from repeatedly sampling the population (created under a random chance model), directing them to determine if a sample is surprising (unlikely) or not surprising. Cobb argues for such an approach, suggesting that, “randomization-based inference makes a direct connection between data production and the logic of inference that deserves to be at the core of introductory statistics” (p. 1). In addition, he suggests that students can easily grasp the models and interpret the results. However, Cobb’s argument that students can easily create models, run simulations and interpret results is a conjecture that needs to be empirically tested.

There is a growing body of research that reports on the impact of technology on the development of students’ statistical thinking (e.g., Ben-Zvi & Friedlander, 1997; Fitzallen & Watson, 2010; Maxara & Biehler, 2006; Saldanha & Thompson, 2003). In many of these studies, researchers either used TinkerPlots™ or Fathom®, dynamic statistical software designed specifically for teaching statistics. For example, Fitzallen and Watson used TinkerPlots™ with middle school students and found that the technology helped facilitate students’ ability to represent data, create data summaries, and make informal inferences.

Maxara and Biehler studied college students' reasoning while creating models and running simulations with Fathom. They noted that students did experience difficulty in modeling statistical problems and that certain probabilistic misconceptions continued to exist despite the outcome of a simulation showing evidence to the contrary. Saldanha and Thompson documented that high school students could not easily envision the process of repeated sampling that underlies the construction of sampling distributions. While this body of research reveals some of the affordances of technology as a learning tool for teaching statistical concepts, more research is needed to fully support Cobb's (2007) conjecture of the pedagogical value of using computers to model and simulate data when teaching informal inference. We need to understand how students might construct models, run simulations and interpret results, and what kinds of challenges students might have within such a pedagogical approach. The research reported here describes some of the successes and challenges students in our study faced when modeling and conducting simulations to answer statistical problems.

19.3 Methods

Data was collected in an introductory statistics course at a large urban university in the Northwest region of the United States. The first author was the classroom instructor and the third author assisted with classroom activities and data collection during the quarter. This particular introductory statistics course was designed for students prior to entering the traditional introductory statistics sequence (descriptive statistics, probability, inferential statistics). Students enrolled in this course as a prerequisite for the traditional sequence or to satisfy the required math elective needed to graduate. A total of 16 students enrolled in the course and all students consented to be participants in the study. All 16 students identified themselves as poor math students and expressed low confidence in their abilities.

The first author implemented the CATALST curriculum materials (Garfield et al., 2012) with some minor modifications. The CATALST curriculum consists of three units and each unit begins with a modeling eliciting activity (MEA, see Lesh et al., 2000). Following each MEA, there are several activities in each unit that guide students through key ideas raised in the MEA (e.g., randomness, chance/null model, informal inference based on a single population, p-value).

Data collection consisted of all student work on in-class activities, task-based semistructured interviews, and student assessment items. Students completed the MOST assessment (see Garfield et al., 2012) at the end of the course as part of their final exam. This chapter focuses on the results of the MOST assessment data for two primary reasons: (1) the final outcomes of student work on activities do not represent the range of reasoning we observed in the MOST assessment because by the end of each class period groups had discussed ideas and often the class discussions led students to consensus on their final write-up for each class activity; and, (2) the MOST assessment provides information about where students were after completing a 10-week experimental course focused on MEA, modeling, simulation, and robust use of TinkerPlots™ software.

The MOST assessment consists of four problems, and this chapter reports on results from the first two problems. The first problem, the Facebook Task, investigated changes in relationship status and particular days of the week when relationship status changed. The task stated that in a random sample of 50 breakups reported on Facebook, 20 % occurred on Monday. Students were then asked to explain how they would determine if the result found by the researchers was surprising under the assumption that there is no difference in the chance for a relationship breakup among the 7 days of the week (null model). The second task, the Music Task, described an experiment where a music teacher plays one of seven notes (A, B, C, D, E, F, or G—no sharps or flats) one at a time, the teacher plays a total of 10 notes and after each note is played a music student must try to select the correct note. Students are then told a music student identified seven notes out of ten correctly and asked if this result would be surprising if this student was merely guessing.

In our reflections on the goals and purposes of the CATALST curriculum, as well as our synthesis of the research literature on modeling, simulations, and informal inference (e.g., Chance et al.'s (2004) work on student thinking about sampling distributions in the context of computer simulations; Cobb's (2007) call for teaching the logic of inference using computer simulations), we concluded that there are four primary phases of reasoning one goes through in the statistical modeling and simulation process using TinkerPlots™ (see Table 19.1). We conjecture that students need to understand each of these phases and the role each phase plays in answering informal inference questions in order to have robust knowledge of the logic of inference and how to use computer technology to model and answer a statistical question.

During the implementation of the CATALST curriculum, the first and third authors observed students thinking through the four phases (Table 19.1) of the statistical modeling and simulation process in different ways, from uncritical attempts at modeling, merely following TinkerPlots™ directions from a prior activity, to statistically meaningful thinking at one or more of the four phases, and finally to metacognitive types of thinking about the entire process and critically evaluating alternative models and the results of those models. Based on these observations, and the idea of mode-level reasoning developed by Ben-Zvi and Friedlander (1997) to describe middle school students' statistical thinking in technology-rich environments, we constructed five modes of statistical thinking within the four phases of modeling and simulation (see Fig. 19.1). The mode number corresponds to the number of phases a student provided explicit evidence of statistically meaningful reasoning. For example, a student might express statistically meaningful reasoning about setting up a model for the statistical problem (phase 1), but not be able to create a meaningful representation of one trial of the simulation or multiple trials (phases 2 and 3), and thus would be coded as Mode 1—Phase 1. Alternatively, a student may create a model that is not statistically meaningful and, thus, cannot adequately model the statistical problem (phase 1), but they may be able to construct a reasonable and meaningful representation of one or more trials of the simulation and/or be able to discuss the results in ways that suggest evidence of understanding the null model and the logic of inference (phase 4). Such an example would be coded as Mode 3—Phases 2–4.

Table 19.1 Phases of reasoning in the statistical modeling and simulation process

Phases of the statistical modeling and simulation process (via TinkerPlots™ technology):	
Phase 1	Model set up that allows for an accurate simulation of the experiment (statistical problem). Constructing the TinkerPlots™ sampler to appropriately model the statistical problem and be able to interpret what <i>draw</i> , <i>repeat</i> , and <i>trial</i> represents within the context of the problem
Phase 2	Running a single trial of the experiment (as set up in the model), investigating the outcomes from the single trial, constructing a suitable representation of the outcomes from the single trial, and interpreting the results of a single trial
Phase 3	Understanding what the variable of interest is, “Collecting statistics” on the variable of interest that was created in the single trial from Phase 2, creating the null (or chance) model distribution from the collection of statistics, describing the process that allowed for the construction of the sampling distribution, interpreting the resulting distribution within the context of the problem (e.g., what does the data in the chance model represent?)
Phase 4	Coordinating the actions of Phases 1–3 and drawing conclusions. Comparing the observed sample data from the problem to the null model generated by the simulation and interpreting the results as well as drawing inferences based on the data at hand. An explicit description of what the null model represents and how it allows for inferences based on the observations from the sample

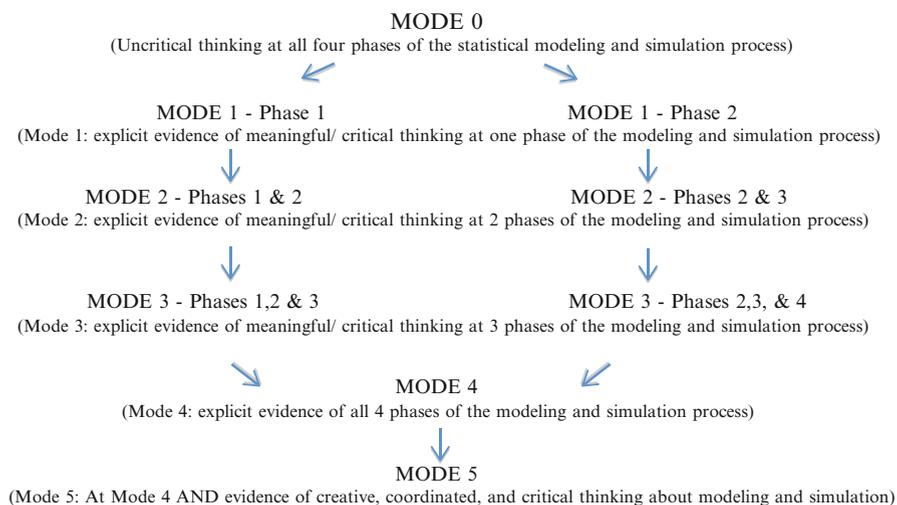


Fig. 19.1 Interaction of mode-level reasoning and the phases of statistical modeling and simulation

19.4 Results

In order to provide a sense of students' statistical development over the course of the quarter, we briefly describe the mode-level reasoning for students at the beginning of the quarter when they first encountered Tinkerplots™ software and were asked to model and simulate statistical problems. The first TinkerPlots™ activity contained three parts: modeling the flip of a coin 10 times and tracking the number of heads, modeling the roll of a die 10 times and tracking the number of three's, and modeling drawing 10 cards from a standard deck and tracking the number of hearts. The activity guided students through setting up each model and then asked them to answer questions about whether or not particular outcomes for each experiment were surprising and why or why not. The first author modified this activity by taking away the directions for modeling drawing 10 cards from a deck and tracking the number of hearts, instead students were asked to construct a model. There were three students who successfully modeled the card task (Mode 1) and two students who could also explain what a single trial of the simulation would look like and the variable of interest, hearts (Mode 2). None of the students knew what a sampling distribution was or had a sense for building a collection of the counts of hearts for many trials of the experiment. For the Facebook and Music tasks (occurring at the end of the term), about half of the students were coded at Mode 3 or 4 and the other half coded at Mode 0 or 1. Data was coded according to the scheme described in the methods section and inter-rater reliability was 97 % for 2 coders. Table 19.2 shows the frequency counts for the Mode level assigned to students for each of the two MOST tasks, as well as the Card activity.

19.4.1 Successes

Most of the class had no conception of the phases of statistical modeling and simulation at the start of the term, but by the end of the term half of the class was coded at Mode 3 or 4. Student descriptions varied, but to be given a Mode 4 code they appeared similar to the following example ("Lena's" write up for the Facebook task):

Table 19.2 Mode-level reasoning for the 16 student participants on the tasks

Mode level	Cards task	Facebook task	Music task
0	13 (81.25 %)	2 (12.5 %)	4 (25 %)
1	1 (6.25 %)	5 (31 %)	4 (25 %)
2	2 (12.5 %)	0	0
3	0	3 (19 %)	4 (25 %)
4	0	6 (37.5 %)	4 (25 %)

To determine whether 20 % of breakups occurring on Monday is due to random chance instead of another factor, I need to construct a null model that assumes there is no difference between the days (**Phase 4**). I started with a mixer sampler with seven items, each labeled for a day of the week. I set the draw value to 1, because only one breakup is happening for each individual, and the repeat value is set to 50, because we are looking at 50 breakups... (**Phase 1**). After running one trial (**Phase 2**), “collect statistic” was used on the percentage of Monday breakups. I plotted the results.... In this case, the p-value is approximately .16, because 16 % of the data is 20 % or higher. This means that it’s 16 % likely to see 20 % or more breakups happening on Monday (**Phase 3**). From this simulation I wouldn’t say that the observed data in the random sample is surprising to see in a chance model (**Phase 4**).

Another example of a Mode 4 response is “Brandon’s” write up for the Music task (his TinkerPlots™ results are shown in Fig. 19.2):

To determine how surprising this result (correctly identifying 7 notes out of 10) is and whether it is strong evidence that the student wasn’t just guessing the notes, I will use TP to simulate a null model where a student’s guessing the notes at random. There will be an equal probability for the simulated student to guess each of the 7 notes (**Phase 4**). Then I will conduct 1000 trials and see how many times the simulated student guessed 7 or more of the notes correctly (**Phase 3**). I set up a sampler with two mixer devices each with seven balls. ...I set the repeat to 10 to simulate 10 notes being played and 10 guesses being made (**Phase 1**). ... Because I collected 1000 statistics and there were none with 7 correct (**Phase 3**), I know that the p-value for guessing correctly 7 times out of 10 is less than 0.1 %. This is strong evidence against the null model (**Phase 4**).

The write-ups of these two students provide evidence that they could identify an appropriate null model and subsequent simulation to create a sampling distribution, which provided statistical evidence to answer the problems they were given.

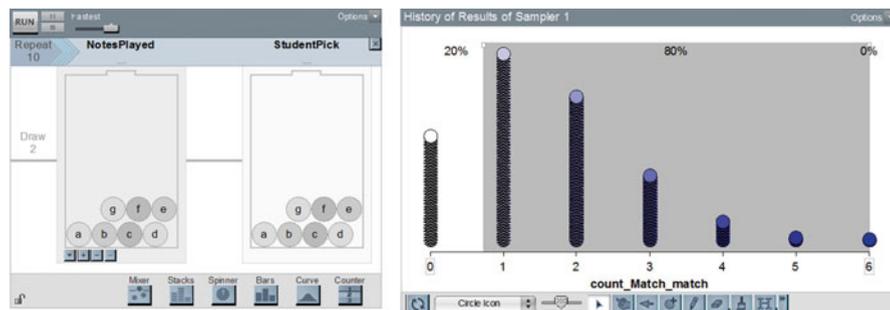


Fig. 19.2 “Brandon’s” model and simulation results for the Music task

19.4.2 Challenges

There were two challenges with the modeling and simulation approach that caused problems for many of the students in this class: (1) the inability to create statistically appropriate representations of the simulation and interpret results based on those representations; and, (2) the inability to set up statistically appropriate models.

In the Facebook task, two students created a valid model, but their choice of representation for one or more trials was not appropriate. Figure 19.3 shows an example of one student’s model and her representation of one trial of the simulation.

This student received a Mode 1—Phase 1 code. Her model (phase 1) is statistically meaningful, as randomly assigning each person to either breakup on Monday or on another day of the week in such a way that the percentage of breakups for Monday (14 %) represents what one would expect under a null model that Monday is just as

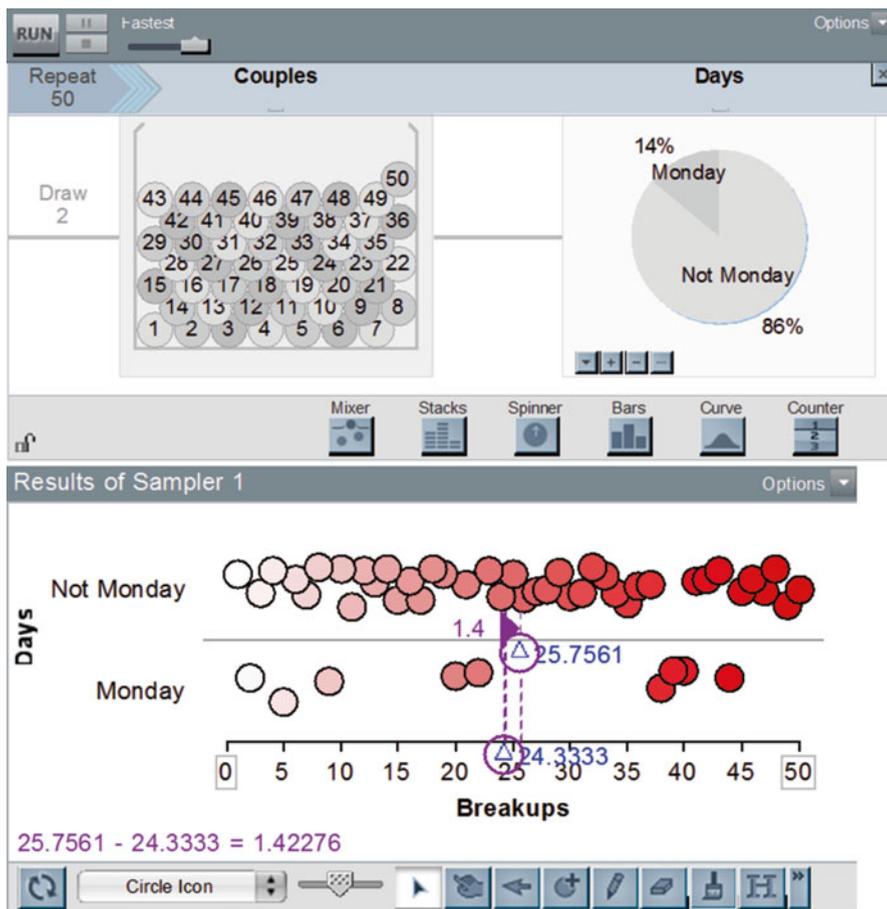


Fig. 19.3 Example of a correct model, but statistically meaningless choice of representation

likely as any other day for a couple to breakup. However, when she creates a representation of her model she treats the number of couples breaking up (on the x-axis) as continuous data, rather than the particular number a couple was assigned. She also graphs Monday breakups with non-Monday breakups and then calculates the mean difference. She went on to collect statistics on the mean difference and create a sampling distribution of mean differences, despite the fact that the mean value is senseless here. Before students took this assessment they were doing activities that required them to find mean differences in flight delay time for two different airlines. It is likely that the two students who completed their assessment in this way were copying the representations from this prior model, rather than making statistically meaningful choices of representations with the data on breakups.

In other instances, students struggled to set up an appropriate model. Figure 19.4 shows an example of a student who received a Mode 3—Phases 2–4 code.

While he included the 7 days of the week in his model, he attached that bin to a spinner that modeled 20 % breakup and 80 % stay-together. This is problematic because it does not represent a model that will simulate the null distribution (e.g., a random chance model), but rather simulates the 20 % breakups mentioned in the observed sample data. He was able to create statistically meaningful representations of one or more trials of the experiment. For instance, he displays a graph of

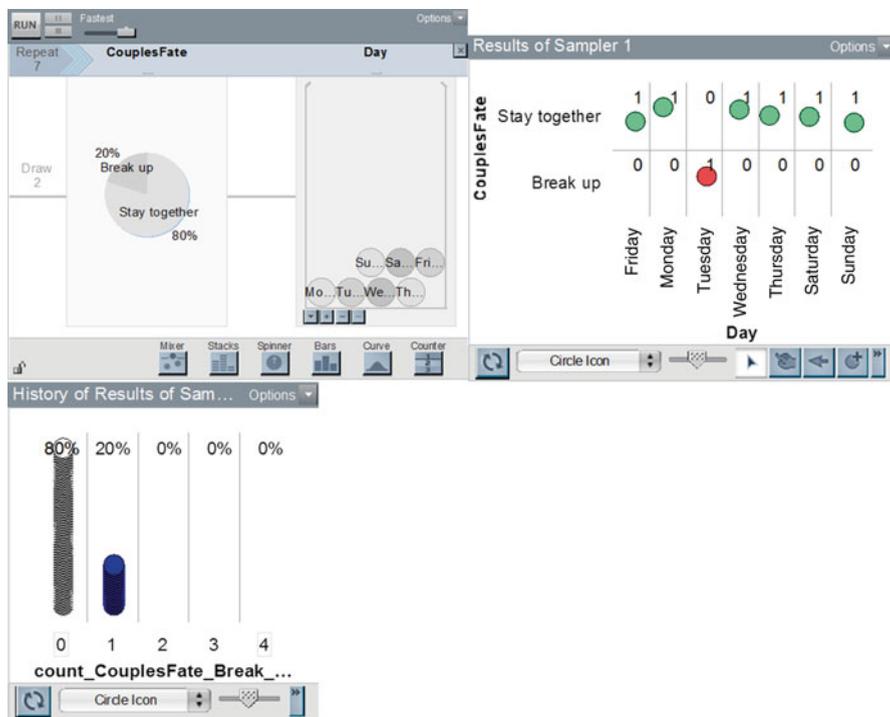


Fig. 19.4 Example of a student’s incorrect model, but statistically meaningful representations



Fig. 19.5 Example of an incorrect application of a 50/50 model

one trial of seven and looks at the counts for breakups and staying together over the days of the week. He continued to collect statistics on the count of breakups on Mondays and interpreted his results in an appropriate way. However, because his model was incorrect he could not correctly answer the statistical problem. This particular issue, *modeling the observed sample data*, was detected during class activities throughout the term.

Another common incorrect model involved student's conflating the idea that there are two outcomes for the music student's guess, correct or incorrect, with the probability that there is a 50 % chance of guessing correctly or incorrectly when in fact there is only a $1/7$ chance of guessing correctly because there are seven notes to choose from (see Fig. 19.5).

Throughout the term, we observed students assigning two outcomes with a 50/50 probability, rather than considering the probability that should be assigned to each outcome.

19.5 Discussion and Conclusions

About a quarter of the students were able to consistently build correct simulation models, run a single trial of the simulation, and interpret that result. Half of the class was not only able to build correct models and run a single trial, but were able to also collect statistics on the correct statistic and draw relevant conclusions about the statistical question. Given the weak background and low confidence of these students, and the goal of having students use technology to draw appropriate conclusions about statistical problems, these results were a huge success with regard to student learning. We saw an increase in statistical reasoning from all students over the course of the quarter. In addition, a few students transcended our expectations and were able to use creative statistical reasoning to answer more challenging questions.

However, we also observed challenges for students using this curriculum. One of the biggest challenges students faced was setting up appropriate simulation models. We observed 50/50 models where they did not apply because students equated two outcomes (such as correct and incorrect on a test) as *always* having equal probabilities. We also observed students modeling the observed data provided in the problem, rather than a null model based on random chance. Furthermore, our students often struggled with the notion of collecting statistics because they thought increasing the sample size or the number of trials would be enough to make conclusions about the problem. These results suggest a need to focus on how technology can be best used to support student learning to set up models, run simulations, and interpret results.

There is huge potential for future research in the area of technology in introductory statistics courses. Curriculum supported by the use of dynamic statistics software will become more prevalent in introductory statistics classes and beyond. Following students as they navigate traditional introductory statistics courses versus courses using technology may provide deeper insight into how students think about statistical concepts and ideas in one context versus another. In addition, long-term studies tracking students who go on to take more advanced statistics may provide insight into developmental pathways over time.

References

- 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. Proceedings of the 1996 IASE Round Table Conference* (pp. 45–55). University of Granada, Spain, 23–27 July, 1996. International Statistical Institute Voorburg, The Netherlands.
- Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Teaching Statistics, 1*(1), Retrieved from <http://escholarship.org>
- Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295–323). Dordrecht, The Netherlands: Kluwer.
- Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1). Retrieved from <http://escholarship.org>
- Fitzallen, N., & Watson, J. (2010). Developing statistical reasoning facilitated by TinkerPlots. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010)*, Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.
- Garfield, J. B., & Ben-Zvi, D. (2009). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.
- Garfield, J., delMas, B., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM Mathematics Education, 44*, 883–898.
- Lesh, R., et al. (2000). Principles for developing thought revealing activities for students and teachers. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Maxara, C., & Biehler, R. (2006). Students' probabilistic simulation and modelling competence after a computer-intensive elementary course in statistics and probability. In A. Rossman, & B. Chance (Eds.), *Working cooperatively in statistics education. Proceedings of the Seventh*

- International Conference on Teaching Statistics (ICOTS7, July 2006)*, Salvador, Brazil. Voorburg, The Netherlands: International Statistical Institute.
- Morgan, K. L. (2011). *Using simulation methods to introduce inference*. Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) webinar. Retrieved from <http://www.causeweb.org/webinar/teaching/>
- Nolan, D., & Lang, D. T. (2010). Computing in the statistics curricula. *The American Statistician*, 64(2), 97–107.
- Rubin, A., Bruce, B., & Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 314–319). Voorburg, The Netherlands: International Statistical Institute.
- Saldanha, L., & Thompson, P. (2003). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257–270.

Chapter 20

Students' Understanding of Statistical Terms Having Lexical Ambiguity

Soyeon Jung and Jihyun Hwang

20.1 Introduction

Understanding terms is very important to learn mathematics. Lemke (1990) stated unfamiliar presentation makes it hard for students to find the lecture interesting or valuable. In addition, using specialized language in science is essential to every goal of science education. However, most students do not succeed in achieving their goals. Therefore, science seems more difficult for students than it is. Learning statistics is not different than learning science. There are many specialized terms in the statistic classroom.

Konold (1995) claimed that adults have strongly held intuitions and theories about probability and statistics. However, in many cases, the ideas are at odds with accepted theory. Therefore, the ideas are problematic when teaching statistics. Also, he investigated students' understanding about a lot of concepts in probability and statistics. He found three characteristics in the students' understanding. First, students have theories or intuitions about probability and statistics before an instruction, many of which are opposite with conventional thinking. Second, students' theories are difficult to change. Last, a student can hold multiple and often opposite beliefs about a particular situation.

Shuard and Rothery (1984) said lexical words which have a similar meaning in mathematics and everyday language. Also, Barwell (2005) defined the same or similar words expressing two or more different meanings as the words having lexical ambiguity. Durkin and Shire (1991) classified different types of lexical ambiguity in

S. Jung (✉)
University of Florida, Gwanak-gu, South Korea
e-mail: mathsoyeon@gmail.com

J. Hwang
University of Iowa, Gwanak-gu, South Korea
e-mail: jihyun-hwang@uiowa.edu

Table 20.1 Category by Durkin

Homonymy	The property of some words that share the same form but have different meanings
Polysemy	The property of some words that have two or more different but related meanings
Homophony	The phenomenon where two different words have the same pronunciation
Shifts of application	Occasions where the same sense can be considered from different perspectives

a mathematical context, as they have been identified by linguistics and their categorization is not limited to mathematics. Table 20.1 shows four main types of words.

Barwell (2005) conducted research on how ambiguity is presented in mathematical classroom discourse. He said that ambiguity can be seen as a resource for participants. He critically examined the ambiguity in two data. He concluded that the strict separation between formal and informal language is not always effective in a mathematics classroom. Informal language can be used to explore and develop delicate mathematical ideas and to take part in mathematical exercises. Therefore, he said ambiguity plays an important role as a resource for students and teachers, serving as a means of articulating between thinking and discourse.

Kaplan, Fisher, and Rogness (2009) examined the meanings of some statistical words association, average, confidence, random, and spread which are most commonly used by students entering an undergraduate statistics course. The results showed that each of these words were problematic for different reasons. In the case of average and spread, both had a variety of meanings for the students entering an introduction to statistics class. It was problematic that there were many common uses for the words, but also these uses were not consonant with the statistical uses of the words. Association, random, and confidence were fairly convergent in terms of common usage from the student perspective. They claimed students make incorrect associations between the words they know and the similar words that have different meanings in statistics from the common usage definitions. Therefore, they concluded that linking the every day and statistical meanings of these words should be done with care and statistics teachers should consider the linking.

Lexical ambiguity makes it hard for students to accept terms as the mathematical ones. The impact of linguistic ambiguity on mathematics learning has been studied variously. Although Kaplan has studies about the statistical words having lexical ambiguity, the ambiguity corresponds to the English language. Because there are a variety of languages, the kind of ambiguity varies in each language. This chapter set out to examine how the students understand Korean statistical terms. The methods can apply in any language though the exemplar words may be different in each language.

The mathematical meanings of many terms in Korea are different from their everyday meanings. In this respect, students easily get confused and have significant difficulties in learning these terms. Therefore, mathematics educators' important problems are finding the terms that have lexical ambiguity and identifying how students understand these terms. In this chapter, it is discussed how college students

accept each of five statistical terms which have lexical ambiguity in Korean secondary school mathematics. The main sample is 20 college students whose major is mathematics education. They have not taken any subjects related to statistics in college. The research questions are:

What do college students whose major is mathematics education know about the variance, variable, frequency, class, and sample?

How much does lexical ambiguity affect the understanding of the variance, variable, frequency, class, and sample?

20.2 Methods

To make a determination, qualitative data were collected by surveys and interviews. The data provided what students understand about each statistical term to check if the daily meanings of statistical terms interfere with their learning of statistics. Durkin and Shire (1991) criteria was used to classify the statistical terms. The criteria gave characteristics of terms to make a framework. After the surveys, two coders analysed the students' answers based on the Kaplan et al. (2009) tool. Kaplan et al. (2009) researched American students on the lexical ambiguity. The tool is easy to use and appropriate to apply in Korean terms.

I made a list of statistical terms used in secondary education that were considered to have lexical ambiguity because there was no study on lexical ambiguity in Korea. The meaning of each word is from Chinese. Table 20.2 is a list of the terms with Korean expressions, phonetic symbols, Chinese expressions, and statistical meanings in English.

I chose the five terms which daily meanings are familiar to students. Table 20.3 shows the five chosen terms that are categorized by Durkin and Shire (1991) criteria.

The sample used during the research was 21st year undergraduate students in the department of Mathematics Education at Seoul National University. The 1st year students had not taken any college statistics courses. They were remarkably

Table 20.2 Terms which were considered to have lexical ambiguity

Korean [phonetic symbol]	Chinese	Statistical term in English
분산 [Bun-San]	分散	Variable
변인 [Byeon-In]	變因	Variance
도수 [Do-Su]	度數	Frequency
계급 [Gye-Geup]	階級	Class
편차 [Pyeon-Cha]	偏差	Deviation
표본 [Pyo-Bon]	標本	Sample
상관 [Sang-Gwan]	相關	Correlation
신뢰 [Sin-Roi]	信賴	Confidence

Table 20.3 Five terms familiar to students

Term	Category	Statistical meaning	Daily meaning
Bun-San (分散)	Polysemy	Variance	Scatteration
Byeon-In (變因)	Polysemy	Variable	A factor that makes others to change Stranger
Do-Su (度數)	Homonymy	Frequency	The number of repetition The figure that represents the size of something
Gye-Geup (階級)	Homonymy	Class	Rank
Pyo-Bon (標本)	Polysemy	Sample	Exemplary behaviour or person

mathematical oriented compared to other students their age. In addition, they were not only college students but also pre-service teachers who will teach statistics in secondary schools.

I conducted the surveys twice. The first survey was done on March 16, 2012. The second survey was done on March 23, 2012. All subjects had 1 h to complete the survey after calculus class. The students were given a questionnaire asking five sets of two questions. The first set asked two questions about *Byeon-In*. The students could write more than one answer for each question.

- (a) Define the word *Byeon-In*.
- (b) Build sentences using *Byeon-In*. The same questions were repeated in each set.

Six students were chosen for interviews because their answers of the first survey were confusing. The interviews were done on March 19, 2012. Each interview took about 20 min.

I first started finding the daily meanings of terms based on *the Standard Korean Language Dictionary (2008)*. Statistical meanings of five terms were referred to *1st grade middle school mathematics textbook (Jung, Kim, Oh, & Lim, 2011)*, *3rd grade middle school mathematics textbook (Yoon et al., 2011)*, *Integration and Statistics (Lee et al., 2011)*, *the Basic of Calculus and Statistics (Lee et al., 2011)*. I made the first framework using these materials. The six students' answers were coded according to the framework. Two coders analysed the original data independently. After that, I compared each result and discussed with the other coder.

20.3 Results

This chapter set out to examine how the students understand statistical terms. The results of the data analysis for each term were divided into three sections in alphabetical order. First, the agreement ratio between two coders is shown. The second part is the results of data analysis for the first question related to the definition of terms. Next is a table describing the definition given by the students. This part is the core of this chapter. Last, the second question studied the students' sentences using each term. There were some disparities between the definition and the meaning in the sentences.

Table 20.4 Agreement between coders

Term	Bun-San	Byeon-In	Do-Su	Gye-Geup	Pyo-Bon
Ratio	.90	.95	.75	.95	.75

20.3.1 Coding

After creating coding categories, two researchers independently coded all responses. Table 20.4 indicates the agreement ratio of initial coding. Two researchers discussed coding categories and coded data to research an agreement between the two researchers. After analysis of the first survey, I found two unexpected problems. First, some answers did not have correspondence between a written definition and the written sentence.

Therefore, I conducted data analysis based on only the answers of the first question, define the word. Second, I needed to subdivide our framework because some of the students answered a part of one meaning in the dictionary or combination of two different meanings. Thus, I made a final framework and I again conducted coding the data with the final framework.

20.3.1.1 Bun-San

Bun-San means a variation in statistics, whereas scatteration in everyday life. Five students answered the daily meaning such as something scattering, scatteration, and a degree of dispersion. One student had no answer, and the other students defined Bun-San as the statistical meaning. The most frequent answer (45 % of all students) defined Bun-San as a degree of distance from the representative value, but it was an incorrect answer. During the interview, the students described the variance in relationship to the width of the normal distribution. Therefore, it made students to connect a variation with the distance from the representative value. "Sum of deviation" and "a degree of dispersion of data in statistics" are examples that indicated incorrect statistical meanings. However, only five students wrote correct answers. Among them, two students mentioned just the equation without words, $\sum(x_i - m)^2 p_i$. Table 20.5 shows the replies of subjects to the definition of Bun-San.

Two students wrote the equation which means a variation. They defined the meaning of Bun-San without a word. During the interview, the students could calculate Bun-San though they did not know what Bun-San was.

20.3.1.2 Byeon-In

Byeon-In has three meanings: (1) a variable in statistics, (2) a factor that makes others change, or (3) a stranger in daily life. Fifty-five percent of the students are characterized by the general meanings, such as a changing factor, a factor that makes something change. All daily meanings written by students included "factor" and "changing"

Table 20.5 Student definitions of Bun-San

		Definition	Number of Answers
Daily Meaning		Scatteration/Something scattering/Scatter	4
		Degree of dispersion	1
Statistical Meaning	Incorrect	Sum of deviation	1
		Degree of dispersion of data or variation in statistics	5
		Degree of distance from the representative value	9
	Correct	$\sum(x_i-m)^2p_i$	2
		Square of the standard deviation	1
		Mean of the square of the deviations	2
		No answer	1
Total			26

Table 20.6 Student definitions of Byeon-In

		Definition	Number of answers
Daily Meaning		Changing factor	5
		Factor that has a possibility to change	2
		Factor that can make others to change/A factor has an effect on others	4
Experimental Meaning		All data related to changing quantities in experiments	1
		Factor that has an influence on something or a results in experiments	2
		Operated factor in experiences	2
		Factor changing under some conditions	1
		Influenced or influencing factor in experiments	2
		Specific setting that can be assigned to subjects	1
Total			20

simultaneously. “Changing factor” is the most popular answer in daily meaning answers. Five students chose a changing factor as the definition of Byeon-In. On the other hand, 45 % of the students defined Byeon-In in experimental terms. The answers classified into experimental terms were characterized by the statement, “in experience”. Also, they used academic terms such as operated, independent, dependent, and control. Similar to the daily meaning, all statistical answers were also connected to a factor. Although half of the students gave answers in science, there was no correct definition of Byeon-In. Table 20.6 shows the students’ answers to the definition of Byeon-In.

20.3.1.3 Do-Su

Do-Su signifies a frequency in statistics. Do-Su is the number of repetition or the figure that represents the size of something in everyday life. Fifty percent of the students answered daily meaning. The figure representing the size of something was

Table 20.7 Student definitions of Do-Su

		Definition	Number of answers
Daily Meaning		Number of repetitions	5
		Figure that represents the size of something such as angle, temperature, and light intensity.	8
Statistical Meaning	Incorrect	Ratio of data under a specific condition	2
		Frequency in the frequency table	2
		Number for each variable	5
		Figure used to inform distribution in statistics	1
		Number for the set of similar samples	1
	Correct	Number of instances in which a variable takes each of its possible values.	4
		No answer	1
Total			29

the most frequently written (40 % of the students) definition among students' replies. 20 % of the students wrote only daily meaning. It means that 80 % of the students showed statistical meanings, and 30 % of the students gave both daily meaning and statistical meaning. Remarkably, most of the students gave the correct answer. Four students defined Do-Su as the correct answer, "the number of instance in which a variable takes each of its possible values".

Among the incorrect statistical meaning, "the number for each variable" was given by five students. Table 20.7 shows the replies of students to the definition of Do-Su. Two students defined Do-Su as "the frequency in the frequency table" without any additional explanation, and six students also made sentences with frequency table. It shows that a frequency table is an important part of learning a frequency. Activities with a frequency table such as analysing a frequency table and making a frequency table brought the strong relationship between a frequency and a frequency table.

20.3.1.4 Gye-Geup

Gye-Geup means a class in statistics, but a rank ordinarily. Seventy-five percent of the students stated one of the daily meanings. Moreover, 70 % of the students gave only one daily meaning. The ratio of students who answered only daily meaning is the highest among the five terms. The most frequent meaning given by 45 % of the students was "people's status". Just seven students gave statistical meanings. However, there was no common characteristic. One student indicated the correct statistical answer. Table 20.8 shows the results of students' answers about the definition of Gye-Geup.

Table 20.8 Student definitions of Gye-Geup

		Definition	Number of answers
Daily Meaning		Criteria for classification	1
		People's rank or social status	9
		Division or rating based on grade or quality	3
		Range	1
		Group included something	1
Statistical Meaning	Incorrect	Regular interval	2
		Set of data with similar level	1
		Standards for dividing variables	1
		Name for characteristics in an arbitrary sample	1
	Correct	Interval to classify or divide variables	1
Total			21

Table 20.9 Student definitions of Pyo-Bon

		Definition	Number of subjects
Daily Meaning		Exemplary behaviour or person	3
		Samples to examine characteristics	1
Experimental meaning		Subjects for experiments	4
Statistical Meaning	Incorrect	Subjects in statistics	4
		Randomly extracted group from the whole group	4
		Part of the population having representativeness	3
		Space or range where arbitrary conduction can be done in statistics	2
		Subject that makes meaningful values	1
	Correct	Extracted part of the population	2
Total			24

20.3.1.5 Pyo-Bon

Pyo-Bon denotes a sample in statistics, whereas usually means exemplary behaviour or person. In addition, Pyo-Bon indicates a subject of experience. Contrast to Gye-Geup, there are only four answers considered as daily meanings such as an exemplary person and a sample to find characteristics. Ninety-five percent of the students answered the statistical meanings. The incorrect statistical meanings answered by students were classified by two main characteristics: randomness and representativeness. Table 20.9 shows that four students defined Pyo-Bon with randomness and three students did with representativeness. Eight answers considered Pyo-Bon as a subject. Among them, four answers are a subject in biology, and the other four

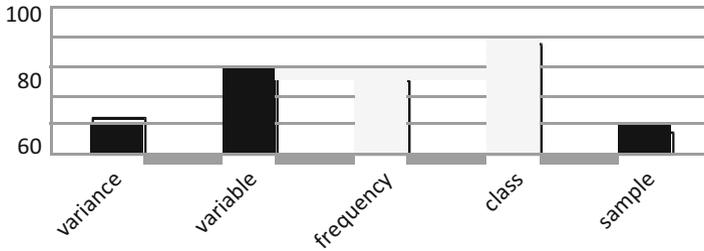


Fig. 20.1 Ratio of Daily Meaning. Note: *Light gray bars* indicate homonymy and *black ones* mean polysemy words

answers are one in statistics. The correct statistical answer was given by only two students.

The students who put an emphasis on randomness did not consider representativeness. It revealed that these students regarded randomness as an essential factor of representativeness. Actually, only random sampling appeared in high school mathematics. Figure 20.1 shows the ratio of daily meaning in the students' answers.

20.3.2 Sentences

The results of analysis for written sentences by the subjects are shown in this section. An example for a daily meaning sentence of Bun-San is "Scatter the people!" Most of the sentences that contained variance, statistical meaning of Bun-San, were related to the scores of tests. For example, "the mean of students' score is 70 and the variation is 5".

Byeon-In in the sentence was not in line with the written definition. For example, one students' definition was "specific setting", but the sentence was "temperature is a control variable in this experiment". All of the students except for one made sentences related to experiments. Remarkably, almost all students mentioned, "operated and control variables in the sentence". The subjects were frequently exposed to the variable in science courses, but they did not learn the variable in secondary school mathematics. For this reason, the students connected the variable with experiments. Though it is the scientific meaning, it is not distinct from the statistical meaning. The control, independent, and dependent variable are also statistics terms. Therefore, it was hard for the students to use the daily meaning in sentences. Daily meaning is involved in only one sentence, "the atmosphere is causing the matter".

In Pyo-Bon, instances for daily meaning were "he is a sample of this job" and "a rat is caged for an experimental the sample". A sentence having Pyo-Bon as a statistical meaning was "it is good to extract sample correctly in statistics". Bun-San, Byeon-In, and Pyo-Bon are in the category of polysemy, the property of some words to have two or more different but related meanings. In this case, the ratio of the

students who answered with the daily meaning is low. Nevertheless, most subjects did not know the correct statistical definition of each term.

In the case of Do-Su, it was distinctive that six students made a sentence using “a frequency table”. For instance, “complete the following frequency table” are a representative sentence given by these students. “The frequency of boys is ten in this statistical data” is an example written by students using frequency alone. The example for usual meaning is “the number of car accidents is increasing every year”.

In the cases of Gye-Geup, an example for daily meanings was “a rank in a military”. When three students made sentences with class, the statistical meaning of Gye-Geup, they used frequency simultaneously. For example, “the frequency of the class about using bicycles is thirty”. Fourteen students gave only the daily meanings and five students did only statistical meanings. In addition, just one student made the correct statistical definition. These remarkable results for Gye-Geup indicate either that the students did not know the definition or that the students could not recognize Gye-Geup as a statistical term. Similar findings appeared in the written sentences by the students. Actually, Gye-Geup appeared very often in everyday life. Activity with a frequency table also confused the definition of Gye-Geup similar to Do-Su. Do-Su and Gye-Geup are in the category of homonymy, the property of some words to share the same form but have different meanings. In this case, the high ratio (50 % for Do-Su, 75 % for Gye-Geup) of the students answered with the daily meanings. Also, some students (30 % for Do-su, 5 % for Gye-Geup) wrote the daily meaning and statistical meaning together. But, these students clearly distinguished the daily meanings from the statistical meanings.

20.4 Discussion

The most frequent answer of Bun-San was “a degree of distance from the representative value”. This is a misconception of a variance. In addition, there were some students who defined Bun-San with an equation. This revealed that some students were able to calculate a variance though they did not know the definition of Bun-San. In the case of Byoen-In, most of the answers were related with experiments. There are two main findings. First, there was a difference between the sentences and the definition given by the students. All sentences except one were related to scientific experiments regardless of their own definition. Second, the results showed that all of the statistical meaning given by the students included “change” or “influence” and “factor”. The figure representing the size of something was the most frequently written (eight students) definition of Do-Su among student replies. A considerable number of students made the definition of Do-Su with a frequency table. Especially, a frequency table appeared in both the definitions and the sentences. Most of the students gave only the daily meanings of Gye-Geup among the five terms. Almost half of the students considered Gye-Geup as a person’s status. Seven students’ statistical answers had no common characteristic. In the case of Pyo-Bon, the number of students who wrote the daily meanings is the least of the five terms. It results from the

fact that there are two Korean words independently, *Sample* and *Pyo-Bon*, although a sample is translated to Pyo-Bon in Korea. Sample is used mainly in daily life, whereas Pyo-Bon is used for experiments or statistics. For this reason, the ratio of the students who gave the daily meaning was lower than any other terms. Moreover, the statistical meanings given by 95 % of the students were divided into two groups. One group was characterized by randomness, and the other group was characterized by representativeness. A random and representative sampling is emphasized in Korean secondary schools. Thus, the students recognized randomness and representativeness as essential characteristics of a sample.

The polysemy words, Bun-San, Byeon-In, and Pyo-Bon, have a relation between a statistical meaning and a daily meaning. In this case, the ambiguity can be seen as a resource for instruction as Barwell (2005) argued. Ambiguity can be important, however, it demands great caution. The subjects gave the various meanings for each term. Considering the reasons of these results, two keys can be found in Konold's research. The first one is the students' intuition related to daily life. A variety of the intuitions had a great influence on the meaning which the students thought diversely. The second is multiple beliefs. The case of Byeon-In which appeared respectively in different courses shows not only that the students can have multiple beliefs but also that the students can have an overwhelming experimental meaning among the multiple beliefs. In addition, the other cases indicated the students' multiple beliefs such as the equations and a frequency table. The students interpreted the statistical words as daily meanings in practice. This results are in line with Kaplan et al. (2009)' idea and inform that her idea similarly established in Korea.

References

- Barwell, R. (2005). Ambiguity in the mathematics classroom. *Language and Education*, 19(2), 118–126.
- Durkin, K., & Shire, B. (1991). Lexical ambiguity in mathematical contexts. In K. Durkin & B. Shire (Eds.), *Language in mathematical education: Research and practice* (pp. 71–84). Philadelphia, PA: Open University Press.
- Jung, G. S., Kim, J. H., Oh, J. R., & Lim, Y. Y. (2011). *1st grade middle school mathematics textbook*. Seoul: Daegyo.
- Kaplan, J. J., Fisher, G. D., & Rogness, T. N. (2009). Lexical ambiguity in statistics: What do students know about the words association, average, confidence, random and spread? *Journal of Statistics Education*, 17(3), 1–19.
- Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, 3(1), 1–11.
- Lee, J. Y., Choi, B. R., Kim, D. J., Seo, J. I., Chun, Y. J., Kim, H. S., et al. (2011). *Integration and statistics*. Seoul, South Korea: Chunjae Education.
- Lee, J. Y., Choi, B. R., Kim, D. J., Seo, J. I., Chun, Y. J., Kim, H. S., et al. (2011). *The basic of calculus and statistics*. Seoul, South Korea: Chunjae Education.
- Lemke, J. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex.
- Shuard, H., & Rothery, A. (Eds.). (1984). *Children reading mathematics*. London: John Murray.
- The National Institute of the Korean Language. (2008). *The standard Korean language dictionary*.
- Yoon, S. S., Kim, H. G., Cho, N. S., Kim, H. Y., Cho, J. M., & Jung, S. Y. (2011). *3rd grade middle school mathematics textbook*. Seoul, South Korea: The Text.

Chapter 21

Student Understanding of Symbols in Introductory Statistics Courses

Hyung Won Kim, Tim Fukawa-Connelly, and Samuel A. Cook

21.1 Research Questions

In the field of mathematics, significant importance was placed upon symbolic representations of communication, teaching, and learning (Arcavi, 1994). In particular, students at introductory level statistics courses have been found to mix up the symbols for statistics and parameters (Mayen, Diaz, & Batanero, 2009), which could hinder them from developing the concepts that such symbols represent. However, our literature search suggests that there have not been any studies published that explore students' understanding of the symbolic system of statistics. Therefore, we investigate the following questions:

- How do students perceive the symbols for mean and standard deviation after a lecture course?
- How does students' symbolic fluency relate to their ability to make sense of more advanced statistical concepts?
- When students have a strong mathematical background, how does that support or inhibit their ability to be successful in developing symbolic reasoning in statistics?

H.W. Kim (✉)

The University of Texas Rio Grande Valley, Edinburg, TX, USA
e-mail: hyung.kim@utrgv.edu

T. Fukawa-Connelly

The University of New Hampshire, Durham, NH, USA
e-mail: Tim.FC@unh.edu

S.A. Cook

Wheelock College, Boston, MA, USA
e-mail: scook@wheelock.edu

Previous research suggests (Mayen et al., 2009) and our results confirm that students find the choices of symbols arbitrary and difficult to associate with related concepts, and that students need particularly strong conceptual and symbolic understandings in order to make conceptual sense of the standard deviation of a sampling distribution. We also found that student understanding of the relation of *statistics* to *parameters* was not robust, and they did not consistently view statistics as variables. We found that many students did consistently look for meaning based upon the symbolic representation of concepts.

21.2 Literature Review

Onto-semiotic research proposes that “representations cannot be understood on their own. An equation or specific formula, a particular graph in a Cartesian system, only acquires meaning as part of a larger system with established meanings and conventions” (Font, Godino, & D’Amore, 2007, p. 6). The implication is that the system of practices is complex in that each one of the different object/representation pairs provides, without segregating the pairs, a subset of the set of practices that are considered to be the meaning of the object (Font et al., 2007). Within the realm of statistics, even when the object under consideration seems relatively simple, such as the mean, there are often multiple symbolic representations used interchangeably. For example, $\bar{x} = (\sum x_i) / n$ may be used without consideration of any other type of representation: graphical, verbal, etc. The relationships between object and representations become even more complex when moving toward a more complex idea, such as the standard deviation of a sample mean. Due to a layering of representations, it is conceivable that the different possible pairs of object/representation convey different meanings of the same object.

For example, when learning the standard error of a sample mean, students are confronted with the simple looking formula: $\sigma_{\bar{x}} = \sigma / \sqrt{n}$. This formula has a seemingly simple explanation: “the population standard deviation of the sample means is given by the population standard deviation divided by the positive square root of the sample size.” In this case, the representation $\sigma_{\bar{x}}$ draws on the agreed-upon symbols for the population standard deviation and the sample mean to communicate the meaning “the population standard deviation of the sample means.” However, it does not give information about how to determine the value. Moreover, the symbol $\sigma_{\bar{x}}$ requires students to be able to make sense of a mixture of previously separate representational systems: those that represent statistics derived from a sample (for example, \bar{x}) and those that represent parameters derived from a population (for example, σ). When given the representation on the right-hand side of the equation, $\sigma_{\bar{x}} = \sigma / \sqrt{n}$, students read a formula that implies they should perform a calculation by mixing the pieces of symbols up from separate representational systems. Most importantly, the right- and left-hand symbols could be interpreted as different meanings of the standard error of the sample mean. So, students are potentially confronted with various possible representations of the same object as described above.

A representation is “something that can be put in place of something different to itself and on the other hand, it has an instrumental value: it permits specific practices to be carried out that, with another type of representation, would not be possible” (Font et al., 2007, p. 7). In this case, the standard error of a sampling distribution, the object $\sigma_{\bar{x}}$, can be understood as a necessary concept that emerged from a system of practices although the choice of symbols can be understood as arbitrary (Hewitt, 1999). It should be considered unique, with a holistic meaning that is agreed upon by the community of practice; however, the concept is expressed by a number of different representations. Hewitt pointed out that for students to communicate with experts, they must memorize the arbitrary elements, the symbols used to represent concepts, and correctly associate them with appropriate understandings of the necessary elements. Each of these object/representation pairs should be understood as encapsulating a different possible set of meanings and enabling different practices.

21.3 Rationale

A literature search suggests that although there have been investigations of students’ understanding of measures of center (Mayen et al., 2009; Watier, Lamontagne, & Chartier, 2011), variation (Peters, 2011; Watson, 2009; Zieffler & Garfield, 2009), and even students’ preconceptions of the terms related to statistics (Kaplan, Fisher, & Rogness, 2009), no one has yet explored student understanding of the symbolic system of statistics. Only one paper has described student errors related to representations of the mean and median (Mayen et al., 2009).

Recently, Shaughnessy called for research into “students’ conceptions of the interrelationships of the aspects of a distribution” (2007, p. 999). But he focused only on the special place of graphs as a tool in statistical thinking, and did not acknowledge the importance of the representational system in which graphs are situated. The research on students’ conceptual understanding of statistical concepts has avoided discussion of the importance of representation; yet, onto-semiotic research claims that descriptions of conceptual understanding are incomplete when pursued only via one or two possible representations of a concept. This study contributes to the growing body of research on student understanding of statistical concepts by describing students’ symbolic fluency and the ways they link concepts and symbols.

21.4 Methods

Data for this study was drawn from eight participants in a mid-sized public university in New England. Two of the participants were in a lower level introductory statistics class and six were from an upper level class. The lower level class was designed to allow first-year students to meet the general education requirement of the university,

and thus is non-calculus based. The upper level class was designed to serve mathematics majors, and thus is calculus-based. The two courses occurred in the same semester. While the curricular organization of the courses in this study conformed to those typically found in a reform-oriented classroom, the instruction itself was essentially traditional. The instructors had almost total responsibility for daily classroom activities and the content was delivered primarily via lecture.

We used a phenomenological approach to collect data, the process of which was conducted in two steps: a survey assessment and a follow-up interview. For the survey, we developed a 14-item assessment. Some of these items were modified from *Assessment Resource Tools for Improving Statistical Thinking*, developed by the faculty members of the University of Minnesota in 2006. The rest of the items were created by our research team. The entire survey is available by request from the first author. The assessment items sought to evaluate student understanding of what the symbols represented and their conceptual understanding primarily via their symbolic representations.

The goal of the interview process was to identify how students' understanding of symbolic representations and their level of symbolic fluency potentially impacted their understanding of certain symbol-oriented concepts. The interview of the two participants from the lower level class was conducted a few days after the survey; the interview of the seven participants from the upper level class was conducted immediately after the survey. Based upon their work on the content survey, the nine students appear to range from low achieving to high achieving in statistics.

Both the survey and the interview were analyzed qualitatively. All interviews were audio-recorded and transcribed. For coding, each utterance was assessed to examine the information it gave about symbolic understandings. Then, within each transcript, we categorized and summarized the utterances that deemed informative understandings by the type of concepts and connections it described with their symbolic understanding. We read within and across categories to develop conclusions. We continually rechecked our conclusions against the data that described the students' proficiencies. In this process, to find out how students' understanding of concepts in descriptive statistics is related with their ability to make symbol sense, parts of the grounded theory approach were blended in.

21.5 Results

Through the data analysis process, we drew three conclusions regarding the pedagogical difficulties that many participants encounter when attempting to reason symbolically in statistics. We also detected that high achieving students face a pedagogical hindrance caused by their academic disposition. A detailed description of the findings is illustrated below.

21.5.1 *Students Find the Choice of Symbols Seemingly Arbitrary and Difficult to Associate with Related Concepts*

According to onto-semiotic research, holding various connections that a concept has with its various expressions is essential for one to internalize the concept. One of the connections is associated with the symbol that typically represents the concept. In introductory statistics courses, many concepts of descriptive statistics are introduced with their associated symbols. The choice of the symbols, however, is somewhat arbitrary and students have difficulty connecting the symbols with the concept that they represent. For example, consider the following claims made by Aaron:

- Aaron: Well, it (μ) is sort of the mean of the whole population. So, it's the big mean, as opposed to the sort of small, local mean (for x).
- Interviewer: Okay. And then, the notation for the smaller one is ...
- Aaron: It seems arbitrary to me. It just seems like they didn't have a good symbol, so they just used x -bar. But it's one of those where I just remember it, because I just had to force myself to memorize that. There's no intuitive connection there, to mean. It's just, someone said that that's what that is. So that's what I remembered it to be.
- ...
- Interviewer: Okay. What about *sigma*, there? What's your understanding of *sigma*?
- Aaron: *Sigma* would be the standard deviation. The *sigmas* actually make more sense. *Sigma*, being the standard deviation, at least there's the relationship, there's s . So, you know, I guess, it's interesting that they used the Greek s for the sort of whole standard deviation, where sort of local, standard deviations have regular, lower case s . But in the case, like, it's more intuitive than x -bar for the observation.

In this example, while Aaron acknowledges the importance of the symbolic connections, he struggles to find such connections. If students do not connect the concepts with their associated symbols in descriptive statistics, they will be hindered from acquiring new concepts about inferential statistics. The items 1, 2, 5, and 6 on the survey were designed to assess students' ability to discern the symbols for statistics from the symbols for parameters. While students' responses on the assessment instrument regarding symbols were 72 % correct overall, they consistently reported, during the interviews, that they struggled to understand the difference between statistics and parameters and to distinguish between the symbols. Consider further, Michael's claims:

I know μ , I just always associate μ with the mean. I wasn't really sure, I don't remember if it was in the population, if it was the mean of the population or the sample, so I just kind of guessed on that one. And, for x , I think I've learned that is also the mean...

He continued,

So, μ would be, like, all the data, and then, sorted, from smallest to largest, and then divided by how many were in the sample.... And then, \bar{x} is, I think \bar{x} is the same, it's just not sorted by smallest to largest. I'm not really sure.

Based on his performance on other items, it appears that Michael knows how to calculate the mean and understands what it implies mathematically. But these are only part of a complete understanding the concept of *mean*. Another aspect of understanding the mean is the ability to pair it with the distinction between sample and population, which Michael was not able to do. Instead, he attributed an incorrect difference of meanings to the two symbols for mean. While he may be able to correctly answer questions that require calculating the mean, the lack of connection may prevent him from acquiring symbolic fluency.

21.5.2 Students Need Particularly Strong Conceptual and Symbolic Understandings in Order to Make Sense of the Standard Deviation of a Sampling Distribution

The concept of the standard deviation of a sampling distribution was determined to be one of the most difficult concepts for students in our survey. When Ian was asked to describe what a particular symbol represents, such as σ / \sqrt{n} , Ian said, "This is the population standard deviation." He continued:

(s is) the standard deviation of our sample. I think we used s in class. I'm not sure. But we used another thing to separate, just like this, our mean in our sample. And so I thought that was what it was.

That is, he understood σ / \sqrt{n} as the sample standard deviation even though the class had used s as the symbol for the sample standard deviation. This implies that he was so unsure in his knowledge that he was willing to believe that a different symbol could be substituted for s and still mean the same thing. Moreover, Ian's responses to the questions were initially definitive; only after further questioning did he admit having any insecurity of his knowledge. Even then, he did not express concern about mixed understandings or possible misattribution of meaning to symbols. We have two more examples that show students' disconnected understanding on the concepts regarding standard deviation. One of them can be seen in the case of Riley as follows:

Interviewer: But what kind of thing can we pull out, from σ and s ? Does s estimate σ ? Or does it estimate any of these things in here?

Riley: s over square root of n estimates σ , I believe.

Also one of our participants, Andrea, was doing very well in her class and had a very firm understanding of statistics and parameters as was shown in the following conversation:

Interviewer: Could you explain what your understanding ... (is about parameters and statistics?)

Andrea: A parameter is just a piece of information about an entire population, and a statistic is a piece of information about the sample, and maybe a statistic is kind of, you use it to kind of guess at the parameter.

Further, when discussing item one, her misconception between sample standard deviation and the standard deviation of a sampling distribution was detected:

Andrea: But I kind of thought these, I had trouble, on my last exam, with, like, the difference between this one and this one. Because, like, I had a problem with—

Interviewer: The sigma over radical N, and S.

Andrea: Yeah. Because I kind of thought, I don't really, I guess I don't know what the difference, because I thought we, in class, we kind of used this to talk about the variability in a sample, but I thought s described the variability in a sample. So, I think I've got those two things kind of confused.

She acknowledges herself that she is confused with the difference between the symbols σ / \sqrt{n} and s . We confirmed this again in the following part of the interview on item 14:

Interviewer: And what is over radical, then? What's the place for that? Why do we ever consider over radical?

Andrea: Well, maybe I, what I thought, maybe, was that, sometimes you know what the, *maybe you know what is, but you don't know what that (σ / \sqrt{n}) is, and you use over radical—*

Interviewer: You mean, we know, we don't know?

Andrea: Maybe, if you do know, I don't know in what situation you would know but you wouldn't know. But maybe you can use this to estimate that one?

Interviewer: You can use over radical to estimate?

Andrea: I don't really know what I'm talking about. [LAUGHTER] But that's my best guess.

After Andrea understood the meaning of "... is an estimator of ...," she made a comment (in bold above) to imply that over radical estimates. One way to explain this misunderstanding is to realize that students are trained to distinguish statistics from parameters through in-class learning. Once students establish the distinction, they habitually try to discern statistics from parameters; yet their work shows that they admit to struggling in doing this. It should be noted that the expression σ / \sqrt{n} has a great potential to confuse new learners because the symbol σ represents a *population* standard deviation, but the process of dividing by radical n is associated with a *sample*. Students can be easily confused as to what σ / \sqrt{n} is associated with because they are trained to distinguish samples from population in order to be able to distinguish statistics from parameters.

21.5.3 *Students Had Difficulty Viewing Statistics as a Variable*

One of the items was designed to find out if students were able to view statistics as variables and parameters as fixed constants. This skill is an essential aspect of understanding the relationship between statistics and parameters and lays the groundwork for understanding the sampling distribution. We found that all eight students had difficulty holding this view. For example, Michael said,

I think a statistic is a calculated value, and a parameter is a, like a, it would be like a boundary that satisfies a value. S , so, I think would be a, I think would be a parameter, because is the statistic. Its [measured estimator?]

Also, Brian said, “because it () is representative of standard deviation. I guess that varies, but—.” When he was asked for the question from interviewer, “Have you thought of \bar{x} as a variable before?,” his answer was “No. I thought it’s more just a sample, as a value that you give to a particular group.” Another example is from Ian. He said, “I didn’t understand that at all. I didn’t know what we were looking at as, what was changing and what wasn’t changing.” However, with some guidance during the interview, some students were able to understand how a statistic could be viewed as a variable. For example, Andrea said,

Well, I guess, I really don’t know, but I guess, my guess would be that, maybe, it would be x and, because maybe μ and σ don’t vary, because they, I don’t feel like I’m interpreting this question correctly, but I think that would be my guess, because maybe μ and σ don’t vary.

These examples imply that without interruption students’ understanding of statistics as a variable was minimal or nonexistent.

21.5.4 *Mathematically Strong Students Experienced Special Kinds of Struggles in Learning Statistics*

One of the research questions was to identify how students with a strong mathematical background develop symbolic reasoning in statistics. Thus, we designed three items (4, 8, and 11) in the survey to evaluate students’ reasoning level of mathematical concepts. Some participants showed strength on the algebraic and probabilistic reasoning that underlies statistical formulas. This strength was first detected via the survey and was confirmed during the interview process. For example, with the three items in the survey, while the average achievement rate of all eight participants for those three items was 61 %, Ian had 100 % and Jen 89 %. Especially, Ian proved to have a firm understanding of the concepts focused on in the three items during the interview. For example, item 8.a in the survey asked:

In a university, 75 % of the students are male and 25 % are female. 5 % of the male students and 15 % of female own a car. For each statement, determine whether it is true?

(a) *We can conclude that 20 % of the students in the university own a car.*

During the interview, he claimed, without doing the real calculation, “I would say it’s between 5 and 15. Probably around 7 %?” Not only was he one of the few students who could correctly describe both the process and concept of a weighted average, but also he was able to give an approximation of the average using the four numbers shown in the question. Ian further proved his mathematical strength with his academic record showing high grades in multiple advanced undergraduate mathematics courses.

One of the characteristics that students with this disposition had was symbolic fluency. Ian, in discussing item 11, claimed:

The center would still be zero. But the standard deviation would be, because you forgot to divide. Because if you divided, if you do the shift first, by μ , you’re centering it at zero. But if you divide x by μ first, then subtract μ , your center would actually be [UNINTELLIGIBLE], because you’re going to decrease your center when you divide by μ , and then you’re going to shift it the original shift. (*)

This remark of Ian’s about z -score shows that he understands the mathematical concepts that underlie the z -score formula. In this remark, it is also evident that Ian has a strong mathematical symbol sense. He was able to describe each of the pieces of the formula in terms of its relationship to function transformation; he described shifts (translations) as happening when subtracting a constant and noted that not dividing by n has no effect on the location of the center. This development of symbolic fluency (or symbol sense?), we suppose, might be the result of Ian’s pedagogical disposition because such a disposition helps students to make sense of the underlying concepts of a statistical expression that use various symbols. Thus this disposition of a student would work as a great pedagogical tool for the student when explanations of statistical expressions are provided to his or her satisfaction.

However, when these mathematically strong students attempt to bring the tools that helped them be successful in K-16 mathematics to their statistics classes, they could feel as though there were different norms for perceiving mathematical concepts in statistics classes because in these classes, contrary to other mathematics classes, it is not common for instructors to provide a complete description of the statistical expressions. As such, participants claimed during the interview that the mathematical concepts were not fully explained in their classes. For example, Ian said, “I feel like we just didn’t get any of the foundational stuff. Like, this is the most lost I’ve ever been in a class.” He further claimed during the discussion of item 7:

And then, there was another question where, you said, like, which of these can be considered variables, or something? Well, I never understood, he never specifically said that, and I never grasped what variables were considered, in stats. So, I guess, when you don’t have that basic, basic stuff, it’s, everything that comes after, you just struggle to try to put pieces together, all at the same time.

This remark not only shows Ian’s frustration that they didn’t learn basic statistical concepts from which they can develop more advanced concepts, but also reflects the conflict with Ian’s pedagogical disposition to seek out an explanation. Now, it seems as though this pedagogical disposition of Ian may have hindered him from developing symbol senses needed to perform well in their class reflects. For example, Ian said, “So, now, I’m questioning myself. This median, capital M , is that the median of the

whole population? Like, can they have the median of the sample? I've never heard that." On one hand, such a deep understanding of statistical expressions and symbolic fluency described above in (*) was the result of the kind of academic disposition that Ian had. But, on the other hand, this academic disposition causes pedagogical conflicts with these students because they feel that the explanations provided are not to their satisfaction.

21.6 Discussion

Students, in introductory statistics courses, often struggle with symbols and making sense of concepts in relation with symbols. In an attempt to elucidate the issue, this chapter addressed the following research questions:

- How do students perceive the symbols for mean and standard deviation after a lecture course?
- How does students' symbolic fluency relate to their ability to make sense of more advanced statistical concepts?
- When students have a strong mathematical background, how does that support or inhibit their ability to be successful in developing symbolic reasoning in statistics?

In investigating the first of the three research questions above, we found that the majority of students made good sense of the basic statistical symbols in descriptive statistics and distinguished the symbols for statistics from those for parameters. However, some students found the choice of symbols seemingly arbitrary and some students had difficulty associating with related concepts and attributed that difficulty to the arbitrary choice of symbols. To alleviate these difficulties, it might be necessary, as a future study, to investigate if it might be necessary that statisticians develop more systematic symbols for novices.

The second research question inquired how students' symbolic fluency relates to their ability to make sense of more advanced statistical concepts. Even though the majority of students were successful in pairing up the symbols for the mean and the standard deviation to the meanings they represent, students, in general, had trouble making sense of more advanced statistical concepts that use those symbols. In particular, it was conspicuous that students did not develop strong conceptual and symbolic understandings in order to make sense of the standard deviation of a sampling distribution. Also, the failure to view statistics as a variable was clearly shown in all eight students. The problem may have less to do with the conceptual challenge of holding that view, but more to do with some students' claim that they never had a chance to think of a statistic as a variable. To help solve this issue, we suggest that instructors give more attention to the concept of the nature of statistics in relation to their corresponding parameters. It remains, as a future study, to find what kind of examples are effective in teaching and learning how statistics vary sample by sample and thus can be treated as a variable in the given context.

The last research question focused on how the academic disposition of mathematically strong students supports or inhibits their ability to be successful in developing symbolic reasoning in statistics. This was shown in Ian's case. He had an academic disposition to seek an explanation of mathematical concepts and showed, during the interview, a strong reasoning ability about the mathematical expressions that use symbols. Our speculation on this matter is that while the academic disposition that mathematically strong students have supports their study in usual mathematics courses, this disposition could cause such students pedagogical conflicts in statistics courses. This phenomenon is attributed to the fact that in traditional statistics lectures, instructors do not provide a complete description of the statistical expressions. In order to mitigate the conflict, it would be necessary for statistics instructors to acknowledge the issue and inform students of the difference between the nature of statistics courses and that of other mathematics courses.

The findings of our chapter now leave us with the following future research questions. First, at the end of an introductory statistics course, students are expected to be able to associate statistical symbols with their accepted statistical meanings and acquire the symbolic fluency. This would lay the foundation for developing a firm understanding of more advanced concepts in descriptive statistics and in the broader domain of inferential statistics. Our study suggested that, without improved practices or more instructional focus, students are likely to continue to create incorrect semi-otic links and experience great difficulty in developing conceptual understanding. This leads to the next question, "what pedagogical approaches help students make better sense of symbol sense?" For example, it would be worth exploring various types of examples with which students can make better sense of symbols.

Second, we found in this study that not providing students with complete explanations of statistical concepts could hinder learning, especially for the students with the academic disposition described above. Thus the following question should be answered: "to what degree should instructors provide the explanations of statistical expressions?" Due to the dual nature of statistical concepts between mathematics and social science, it would be unrealistic to provide complete proofs of statistical expressions in class. Thus, it is important to identify effective pedagogical methods that balance well between the two aspects of the discipline.

References

- Arcavi, A. (1994). Symbol sense: Informal sense-making in formal mathematics. *For the Learning of Mathematics*, 14(3), 24–35.
- Font, V., Godino, J., & D'Amore, B. (2007). An onto-semiotic approach to representation in mathematics education. *For the Learning of Mathematics*, 27(2), 2–7,14.
- Hewitt, D. (1999). Arbitrary and necessary: Part 1 a way of viewing the mathematics curriculum. *For the Learning of Mathematics*, 19(3), 2–9.
- Kaplan, J., Fisher, D., & Rogness, N. (2009). Lexical ambiguity in statistics: How students use and define the words: association, average, confidence, random and spread. *Journal of Statistics Education*, 18(1). <http://www.amstat.org/publications/jse/v18n2/kaplan.pdf>

- Mayen, S., Diaz, C., & Batanero, C. (2009). Students' semiotic conflicts in the concept of median. *Statistics Education Research Journal*, 8(2), 74–93.
- Peters, S. (2011). Robust understanding of statistical variation. *Statistics Education Research Journal*, 10(1), 52–88.
- Shaughnessy, M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second handbook of research on mathematical teaching and learning* (pp. 957–1009). Washington, DC: NCTM.
- Watier, N., Lamontagne, C., & Chartier, S., (2011). What does the mean? *Journal of Statistics Education*, 19(2). <http://www.amstat.org/publications/jse/v19n2/watier.pdf>
- Watson, J. (2009). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal*, 8(1), 32–61.
- Zieffler, A., & Garfield, J. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal*, 8(1), 7.

Chapter 22

Two-Year College Mathematics Instructors' Conceptions of Variation

Monica Dabos

This chapter presents and discusses the results pertaining to three of the questions dealing with conceptions of variation in repeated samples.

Despite the large number of empirical studies investigating students' difficulties with statistics, there are limited studies exploring the level of understanding of statistics instructors. This becomes particularly relevant if we consider a 5-year period report, which showed a 60% increase in enrollment for introductory statistics courses (Kirkman, Lutzer, Maxwell, & Rodi, 2007). Even though the increase in student enrollment is substantial, there is little information about who is preparing those students to succeed in statistics. According to the CBMS report, only 2% of full-time and 2% of part-time instructors at 2-year colleges in the United States have degrees in statistics.

Within statistics, variation is fundamental as it is part of every step of the statistical analysis process. The current literature reveals that students at the K-12 level (Shaughnessy et al., 2004) as well as college students (Meletioun-Mavrotheris & Lee, 2005) struggle to understand the complexities associated with the concept of variation. There seems to be an implicit understanding behind researchers' recommendation that those teaching statistics possess the appropriate understanding of variation. Therefore, the focus of this study was to investigate the extent of understanding of the conceptions of variation held by 2-year college mathematic instructors. A total number of 52 instructors participated in the study from 33 different California community colleges. They responded to a survey designed to explore their conception of variation. There were a total of 16 questions, all of which were previously used in research to investigate students' conceptions of variation. This chapter presents and discusses the results

M. Dabos (✉)
College of the Canyons, Santa Clarita, CA, USA
e-mail: monica.dabos@canyons.edu

pertaining to three of the questions dealing with conceptions of variation in repeated samples.

The results indicate that 2-year college instructors' responses reflect, for the most part, what other studies have found (Reading & Shaughnessy, 2004). For example, it appears that the wording of the question altered the kind of response instructors gave as indicated by the contrast of 12% ($n=6$) of instructors predicting variability in one question, while 77% ($n=40$) predicted variability in another. Additionally, in the short answer question, it was discovered that while different numerical values were reported for repeated sample questions, it did not necessarily indicate an appropriate conception of variation, as some of the predicted values were considered too low, too high, or too broad. Instructors' reasoning also highlights a division between those who readily predicted and justified sample variability and those who gave explanations that do not show that sample variability has been considered. However, the difference does not seem to be highlighted by the instructors' degree or by their statistics teaching experience. This study suggests further research could clarify if the questions utilized in this and prior studies are valid tools for measuring variability. If the questions are proven not to be the problem, then deeper investigation into instructors' knowledge of variability may be required.

References

- Kirkman, E., Lutzer, D., Maxwell, J. W., & Rodi, S. B. (2007). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2005 CBMS Survey*. Washington DC: Conference Board of Mathematical Sciences.
- Meletiου-Mavrotheris, M., & Lee, C. (February, 2005). *Exploring introductory statistics students' understanding of variation in histograms*. Paper presented at the Fourth Congress of ERME, the European Society for Research in Mathematics Education, Sant Feliu de Guixols, Spain.
- 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). Netherlands: Kluwer Academic Publishers.
- Shaughnessy, J. M., Ciancetta, M., & Canada, D. (July, 2004). *Types of student reasoning on sampling tasks*. Paper presented at the 28th Conference of the International Group for the Psychology of Mathematics Education, Bergen, Norway.

Chapter 23

Students' Sense-Making of Graphical Representation in a Basic Statistics Module

Hui Teng Chia

23.1 Introduction

The Singapore mathematics curriculum (MOE, 2001) introduces statistics education in Primary One. An analysis of the syllabi and curriculum materials revealed that solving mathematical problems based on graphical representations is a common theme throughout primary and secondary school statistics education (except for Primary 5 where students are taught the concept of average). Graduating secondary school, students would have encountered graphs such as picture graphs, bar graphs, line graphs, pie charts, histograms, ogives, stem-and-leaf plots, dot plots and box plots. Given that Singapore students in the primary and secondary schools are taught mathematical problem-solving based on these graphs, how well then do post-secondary school students make sense of information presented graphically?

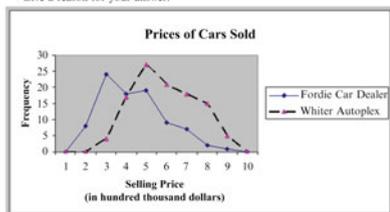
23.2 Methods and Results

Although students were taught mathematical calculations when responding to questions based on graphs, it was surprising that many students chose to use logical reasoning to respond to the task in Fig. 23.1. This made the typical grading of right or wrong solution inappropriate and hence, the SOLO taxonomy was employed to

H.T. Chia (✉)

School of Mathematics and Science, Singapore Polytechnic, Singapore, Singapore
e-mail: chia@sp.edu.sg

1) Two frequency polygons are constructed to compare the prices of cars sold by two car dealers. Which company has an overall higher selling price of cars? Give a reason for your answer.



Answer:

Answer: Shapes of both frequency polygons are about the same but whiter autoplex frequency polygon is more towards the higher selling price.

whiter Autoplex has an overall higher selling price.

~~whiter Autoplex~~ Both have the same selling price. ~~They~~ Both graphs have the same x coordinates for each point, showing that their cars have the same selling price. They only have different number of cars sold.

Fig. 23.1 The task (left panel) and two examples of students' responses (right panel)

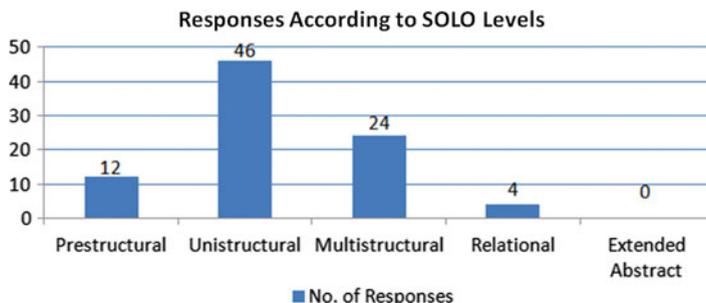


Fig. 23.2 Summary of classification of students' responses

examine the quality of students' responses. Figure 23.2 provides a summary of the classification of students' responses.

One response met three of four characteristics of an extended abstract response and was categorized as the transitional state from relational to extended abstract.

References

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. London: Academic Press.
 Ministry of Education. (2001). Primary Mathematics Syllabus. Retrieved from <http://www.moe.gov.sg/education/syllabuses/sciences/files/maths-primary-2001.pdf>.

Part IV
Technology in Statistics Education

Chapter 24

Bootstrapping Confidence Intervals

Ross Parsonage, Maxine Pfannkuch, Chris J. Wild, and Kate Aloisio

24.1 Introduction

In most undergraduate introductory statistics courses, the conceptual foundations underpinning confidence intervals are the normal distribution, the Central Limit Theorem, and the sampling distribution of estimates. However, research evidence suggests that the theoretical and mathematical procedures involved in statistical inference act as a barrier to student understanding, particularly with regard to hypothesis testing (delMas, R, Garfield, & Chance, 1999; Jones, Lipson, & Phillips, 1994). Although confidence intervals have been taught for years in introductory statistics courses, very little research has been conducted on students' understanding of them (Sotos, Vanhoof, Noortgate, & Onghena, 2007). The sparse research that has been conducted on the understanding of confidence intervals has thrown up a raft of misconceptions. These misconceptions include the thinking that a 95 % confidence interval (for a mean) contains the plausible values for the *sample* mean, covers 95 % of the sample, is the range of individual scores, increases in width with sample size, or is not affected by sample size (Fidler, 2006).

Increasing access to technology, calls to reform the introductory statistics curriculum (Cobb, 2007), and recent changes in statistical practice (Hesterberg, Moore, Monaghan, Clipson, & Epstein, 2009) have led our project team to introduce the randomization (permutation) and bootstrapping methods. The appeal of both the randomization and the bootstrapping methods is that they are logical,

R. Parsonage (✉) • M. Pfannkuch • C.J. Wild
The University of Auckland, Auckland, New Zealand
e-mail: r.parsonage@auckland.ac.nz; m.pfannkuch@auckland.ac.nz; c.wild@auckland.ac.nz

K. Aloisio
Smith College, Northampton, MA, USA
e-mail: kaloisio@gmail.com

accessible, and lend themselves to dynamic visualization processes, which we conjecture may assist students with their understanding of statistical inference. They also do away with the need for distributional assumptions and can be applied to many different situations. In this chapter, we focus on the bootstrap method and our pilot study findings.

24.2 Bootstrap Method Literature

All of the current literature in statistics education on the bootstrap is centered on explaining the method, giving teaching examples, and arguing that the method will give students better access to ideas underpinning inference. There is no research to date on the bootstrap method's effectiveness in improving student learning, on students' reasoning using the bootstrap, or on any learning issues that need to be resolved. Therefore, in this section, we will discuss the bootstrap method and the rationale for using it from a discipline perspective and from an education perspective.

24.2.1 *Breaking with Tradition in Statistics*

In 1979 Brad Efron produced a landmark paper on the bootstrap method that has revolutionized the practice of statistics. Statistical inference is underpinned by the sampling distribution of sample statistics through considering all random samples from the population. Efron's idea was to estimate a sampling distribution from just one sample. By treating this one sample as if it were the population and mimicking the data production process (Hesterberg, 2006), multiple resamples of the same size as the original sample are taken with replacement from this original sample with the statistic being calculated each time. The variation in estimates from resamples from this original sample approximates the variation in estimates that would be obtained if many samples from the population were taken. Since the advent of Efron's paper, well over 1000 papers have been written justifying the theoretical basis for the bootstrap (Efron, 2000). As Efron (2000), p. 1295 noted:

it has taken me a long time to get over the feeling that there is something magically powerful about formulas ... and to start trusting in the efficacy of computer-based methods like the bootstrap for routine calculations.

Since the bootstrap method is capable of generating bootstrap distributions for summary statistics such as medians, quartiles, measures of spread, and correlations, it goes far beyond the scope of classical mathematical methods simple enough to be commonly taught. Cobb (2007) and Efron lament that the bootstrap, which has had a major effect on the practice of statistics is not part of the introductory statistics curriculum. With computing power now available to students, the

time is ripe to introduce students to the bootstrap method and correct the mismatch between statistical practice and the introductory curriculum.

24.2.2 Breaking with Tradition in Statistics Education

In 1976 Simon, Atkinson, and Shevokas used the Monte Carlo method for teaching probability to students. The Monte Carlo method is similar to the bootstrap method in the sense that the probability of an event is estimated through simulation rather than mathematical theory (e.g., Binomial distribution). Simon et al. found that students exposed to the Monte Carlo method did better than students who used conventional methods. They argued that a complete rethinking of how probability was taught was necessary. For students the Monte Carlo method was intuitive, logical, and readily accessible, whereas conventional methods took more time and depended on greater mathematical ability. Their findings and suggestions for changing teaching have remained in abeyance until recently. A concerted effort is now afoot to promote computer-intensive methods in introductory courses (Engel, 2010; Gould, Davis, Patel, & Esfandiari, 2010; Hesterberg, 2006; Holcomb, Chance, Rossman, Tietjen, & Cobb, 2010; Tintle, VandenStoep, Holmes, Quisenberry, & Swanson, 2011; Wood, 2005).

Apart from the fact that bootstrapping is rapidly becoming the preferred way to do statistical inference, there are strong pedagogical arguments for introducing the bootstrap into the curriculum. First, the bootstrap can be used to make the abstract concrete by providing “visual alternatives to classical procedures based on a cookbook of formulas” (Hesterberg, 2006, p. 39). These visual alternatives have the potential to make the concepts and processes underpinning bootstrap inference transparent, more accessible, and connected to physical actions. Student understanding can be enhanced by the addition of visual representations and by encouraging students to generate mental images. Technology also enables students to link multiple representations—visual, symbolic, and numeric—and it facilitates understanding through promoting a visualization approach to learning (Sacristan et al., 2010). Dynamic software can allow students to analyze directly the behavior of a phenomenon, to visualize statistical processes in ways that were not previously possible, such as viewing a process as it develops rather than analyzing it from the end result. Exposure to such processes “can help develop the abilities and intuitive thinking that can enhance powerful mental conceptualizations” (Sacristan et al., 2010, p. 188).

Second, students experience a set of general approaches or a method that applies across a wide variety of situations to tackle problems rather than learning multiple and separate formulas for each situation (Wood, 2005). Such formulas work in special circumstances but the general approach works in most situations and sometimes it is the only option. Third, simulation is currently seen as a teaching aid to improve understanding of inference rather than a replacement for standard methods (Engel,

2010; Hesterberg et al., 2009), since client departments or employers, who are not familiar with more modern practice, may demand traditional methods. However, for the majority of students, who will never need to study analytic methods, simulation methods such as the bootstrap should be promoted as the only method (Wood, 2005). Moreover, these methods coupled with dynamic visualization infrastructure allow access to statistical concepts previously considered too advanced for students, as mastery of algebraic representations is not a prerequisite. As Wood (2005), p. 9 states, simulation approaches such as the bootstrap “offer the promise of liberating statistics from the shackles of the symbolic arguments that many people find so difficult.” Similarly statistics teachers need to liberate themselves from theoretical formula-based teaching and intellectually to accept new ways of practice and thinking.

24.3 Methodology

A collaborative research project team of 33 members is involved in the development of innovative approaches to teaching statistical inference. The team consists of two education researchers, two resource developers, a statistical software conceptual developer, eight university lecturers, 14 secondary school teachers, 5 professional development facilitators, and 1 international advisor. Using design research principles (Hjalmarson & Lesh, 2008), the development process involves two research cycles with four phases: (1) understanding and defining the conceptual foundations of inference; (2) development of learning trajectories, new resource materials, and dynamic visualization software; (3) implementation with Year 13 (last year of high school) and introductory university statistics students; and (4) retrospective analysis followed by modification of teaching materials.

In the first research cycle, a pilot study was conducted with ten students, five from Year 13 and five from university. A one-day teaching session was conducted, with half of the day devoted to the randomization method, and the other half to the bootstrapping method. Data collected were: student pre- and posttests and interviews, videos of teaching implementation, and reflections and observations of the project team. A thematic qualitative data analysis using nVivo was conducted on the student interviews (Braun & Clarke, 2006), while numbers of students who responded to multichoice and true/false questions were recorded. In the second research, cycle data will be collected from about 3000 students.

24.4 Description of Tasks

Our aim is to develop in students an overview of the “big ideas” and purpose of taking random samples and an intuitive grasp of statistical inference by introducing the concepts of confidence intervals via the bootstrapping method. We start students with hands-on experiences of implementing the bootstrap method with a small

number of repetitions and then progress to the next level of the inferential argument using dynamic visual imagery software that closely resembles the hands-on experience but is capable of doing thousands of repetitions. The instruction sequence, which is designed for classes of over 300 students, is now briefly described.

First, the students are introduced to a media article that quotes averages from a survey. Ideas are raised about using a sample to make an inference about the population and whether the average would be the same if someone took a different sample. Since all the estimates reported in the article are uncertain, the question about how to improve on a point estimate is asked. Second, students are shown a population bag of 521 datacards, where one datacard records information about a male student from an actual online survey of introductory statistics students. One of the variables on the datacard is weight. Questions are posed about the shape of the population distribution for the weights and the median weight of the population. A sample of nine cards is taken from the population bag. It is stressed that the situation presented is for teaching them some “big ideas” and that in the real world we do not have the weights of every person and in practice we would take a larger sample. From the nine cards the sample median is calculated. Since this estimate is uncertain, the question about how to find plausible values for the population median is discussed. We then introduce students to the approach Brad Efron used in 1979 and we emphasize ideas such as treating the sample as if it were the population and mimicking the data production. Third, to experience the bootstrapping process the students have nine paper squares with each square representing the weight of the student sampled. In pairs they take a resample of size 9 from the original sample with replacement, plot the data, and record the median. They repeat the process several times. The resample medians are gathered from the class and a plot of the medians represented by vertical lines is built up. Students then suggest what interval of plausible values they would use for the population median.

Fourth, students are reminded that they are trying to estimate the population median and that the bootstrapping process will be automated using dynamic visualizations that we have especially developed. Using the same hands-on data students are introduced to VIT—visual inference tools (see: <http://www.stat.auckland.ac.nz/~wild/VIT>). Development of specialized software was an integral part of the project. Figure 24.1 shows an example of the *first* version of the dynamic visualization tools. The data panels screen shows a tracking feature for sampling with replacement. The top section of the graphics panel plots the observed weight data, the middle section gradually builds up the sampling variability in the resample medians, and the bottom section displays the bootstrap distribution and confidence interval. Each of the four displays is revealed separately and built up dynamically for the students. The lecturer discusses the bootstrapping process relating each part of the display to the hands-on activity. The final display (not shown) shows the confidence interval moving up to rest on the box plot. Note that Fig. 24.1 displays data for a sample of size 36, a later activity given to the students.

Fifth, we ask how successful the bootstrapping process is at capturing the population median. Using our visualization tools, 1000 random samples of size 9 are taken from the population of male weights. For each of these samples, a bootstrap

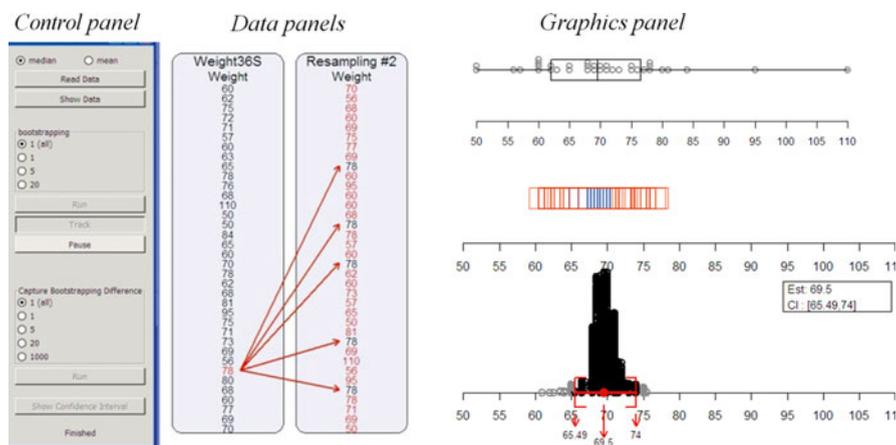


Fig. 24.1 Software panels for performing the bootstrapping process

confidence interval is built (using 1000 resamples) and checked to see whether the population median is in the interval. The software produces numerically and visually the overall success rate for these 1000 bootstrap confidence intervals. Finally, students are introduced to bootstrapping a confidence interval for the population mean and confidence intervals for differences in population medians and means and how to interpret them.

24.5 Pilot Implementation Results

Our main research question for this pilot study was: What issues arise in students' reasoning processes when they experience new methods such as bootstrapping? From the student responses we were specifically interested in: What aspects of the design of the learning trajectories, resources, software, pre- and posttests, and post-task need to be improved before we implement the main study? Data are drawn from written responses and interviews with the students. We highlight three main issues concerning student responses (in the areas of visualizations, "big ideas" and verbalizations) that we learnt about student reasoning and discuss our consequent actions to ameliorate the perceived problem area.

24.5.1 Visualizations

One issue that arose was the number of multiple images or representations for a confidence interval that students were expected to grasp—a band of resample medians, a distribution of resample medians, a numeric interval, a verbalization of the

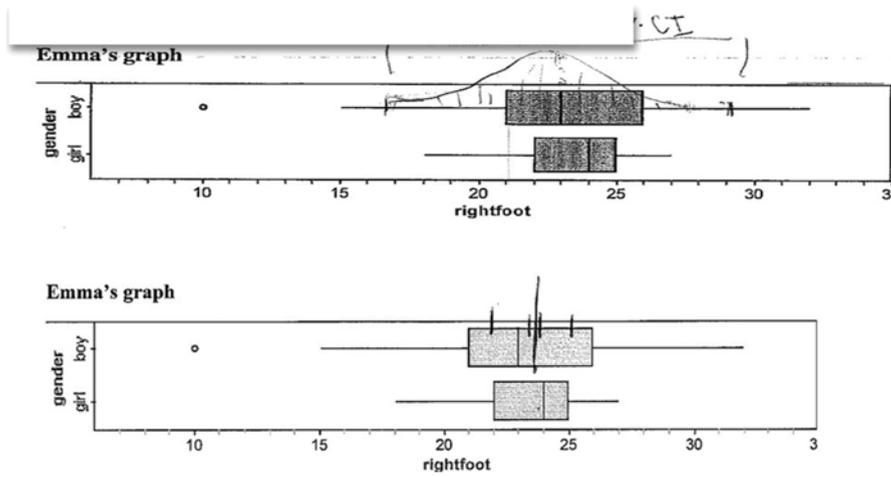


Fig. 24.2 Students' images of a confidence interval for the population median of boys only

interval, and a horizontal line on the original sample. When eight of the students were asked in the posttest interview to draw their image of a confidence interval for the population median for boys only, none of them drew a horizontal line as shown in the software. One student drew the bootstrap distribution (Fig. 24.2), one a partial distribution, three drew marks indicating uncertainty in the median (Fig. 24.2), and the other three put two vertical lines to indicate the boundaries of the confidence interval. When one of them was asked what happened after the confidence interval was calculated on the bottom screen, she said she did not remember. Lack of familiarity with the confidence interval representation image, the fleeting movement of the image to the original box plot, the visual dominance of the bootstrap distribution, and prior knowledge all seemed to have played a role in students missing the final representation of the confidence interval. From responses to other questions, we realized the bootstrap distribution was dominant in their imagery.

Since the bootstrap distribution should just be regarded as a calculating device we decided to change the color of the distribution to a lighter color and incorporate a fade button so that students' attention could be drawn to a more prominent depiction of the confidence interval superimposed on the original data. Furthermore, the resources only gave a numeric representation and an interpretation of the confidence interval, not a plot of the data with the confidence interval represented. These were changed so that the students physically drew the confidence interval on the plot.

24.5.2 "Big Ideas"

In the posttask, pairs of students were interviewed. They were given data on guinea pig survival times in a drug treatment trial. They were asked to teach the interviewer how to obtain the typical survival time of guinea pigs that took this drug. Their

initial reaction was to produce the observed median but when asked if the interviewer could report the observed median as the typical survival time they responded that they needed to produce a confidence interval. The students could easily handle the software, and when questioned about what they doing and why, they were able to discuss the mechanics of the bootstrapping process and describe what each of the visual components represented. Four of the seven students who were asked about the reason for sampling with replacement seemed to be aware that the sample was being treated as if it were the population with comments such as “so that the sample is representative of the population.” The dynamic software visualizations seemed to help students attain the concepts underpinning the bootstrapping process.

What was missing from the student responses were big ideas such as: all estimates are uncertain; if other samples were taken different estimates would be obtained; and to make an inference about a population parameter, a confidence interval as a set of plausible values for the population parameter needed to be constructed. In all of the posttest and task interviews only one student mentioned the word estimate, saying it only once, and none used the word uncertain even though the teacher in the pilot study believed he carefully emphasized this language. We conjecture that the words estimate and uncertain are part of everyday language and therefore did not become part of students’ statistical language or thoughts. Their initial learning seemed to be focused on “what to do to get the answer” rather than the “big ideas” underpinning confidence intervals. Part of the problem may have been the short teaching session but nevertheless we developed the idea of an uncertainty band around an estimate and modified the teaching resources to emphasize and highlight the “big ideas.”

24.5.3 Verbalizations

In the tests and posttask interview transcripts, we noticed students tended not to verbalize terminology such as *sample means* and *population mean* rather they used the word *it* or left the words unsaid. For example, a student stated: “for a large sample size you’re going to get less variability.” To demonstrate further she drew two bell-shaped curves with the one for the larger sample size narrower than the one for the smaller sample size and she did not label or talk about what the *x*-axis was measuring. This practice of not stating what one gets less variability in and not labeling the *x*-axis of the sampling distribution of the statistic was prevalent among the students. When a student constructed a bootstrap confidence interval using the software he gave the following description:

it is a fairly safe bet *it* is somewhere between here ... Rather than just saying *it* will be this many days. It’s more certain that the value you will get from doing another *one* that *it* will fall in between the two rather than saying *it* will be that value.

Another student, who also constructed her own bootstrap confidence interval said:

we observed at the start *it* is 108 and then the confidence interval is 90 to 140.5 so we would be reasonably confident the [*note word omission*] survival time would be between this ... if the guinea pigs took the drug.

All the students seem to understand that the population mean was within the confidence interval with comments such as “where we believe the mean value is of the entire population of guinea pigs.” However, when they were confronted with answering true or false to the following confidence interval statement in the posttest (adapted from delMas, R, Garfield, Ooms, & Chance, 2007): “We believe that it is a fairly safe bet that each cookie for this brand has approximately 18.6–21.3 chocolate chips,” four of them answered true, two answered false with incorrect reasoning such as “you can’t have 0.6 of a chocolate chip,” and four answered false with correct reasoning. Even though the students could clearly state the bootstrap distribution was a distribution of resample medians when looking at and talking about what they were seeing on screen, the idea of a distribution of a statistic seemed to slip from some of their minds when faced with a new situation.

We conjecture that the students’ reluctance to verbalize sample mean and population mean and to label the x -axis for the bootstrap distribution is hindering their understanding and partially preventing them from transitioning from viewing a plot as a distribution of data to a distribution of a statistic. Another reason is that we did not pay sufficient attention to this key transition phase of a new concept of a familiar plot in our learning trajectory. We also noted that our software does not have x -axis labels, for space and visual perception reasons, and our resources have the measure in the title for the plots rather than as labels on the x -axes. We are still thinking about how to address this issue and how to get students to produce a bootstrap distribution visual mental image and what it means when dealing with interpreting a confidence interval in a word-only context.

24.6 Conclusion

The purpose of the pilot study was to detect problems in the pre- and posttests, post-task, learning trajectories, and software before trialing the bootstrap method with over 3000 students. Through interviewing students we identified many issues, some of which are reported in this chapter. We learned that attention to the interplay between verbalizations, language, and the “big ideas” is very important; that visual imagery can go unnoticed; and that imagery has the potential to assist concept development. Unfortunately, it seems that the bootstrap method will not be a panacea for erasing confidence interval misconceptions already identified by other researchers such as students thinking that the confidence interval covers 95 % of the sample (Fidler, 2006).

The strength of our bootstrap software, however, is that the students always see the confidence interval developing visually as part of a distribution, never as a numeric representation alone. Students seem to understand the components of the

software dynamic visualizations and what each represents. Experiencing the visual building up and development of a confidence interval gives students direct access to the behavior of the sampling variability phenomenon (Sacristan et al., 2010). We believe that the bootstrap method coupled with our visual thinking tools allows underpinning concepts such as mimicking the data production process and variation in sample medians to be more accessible and transparent to students. Compared to mathematical confidence interval formulas, we conjecture that students did learn more about statistical inference using the bootstrapping method (cf. Tintle et al., 2011). Questions are currently still being raised about how we can encourage students to make connections between visual, symbolic, and verbal representations of confidence intervals; to make the transition to recognizing a distribution of a statistic rather than as a distribution of data; and to attend to the “big ideas” behind estimations.

Acknowledgements This research is partly funded by a grant from the Teaching and Learning Research Initiative (www.tlri.org.nz).

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1(1), 1–15. Retrieved from <http://escholarship.org/uc/item/6hb3k0nz>
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students’ conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6 (2), 28–58. Retrieved from www.stat.auckland.ac.nz/serj
- delMas, R., Garfield, J., & Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve students’ statistical reasoning. *Journal of Statistics Education*, 7(3). Retrieved from www.amstat.org/publications/jse/v7n3
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26.
- Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistics Association*, 95(452), 1293–1296.
- Engel, J. (2010). On teaching bootstrap confidence intervals. In C. Reading (Ed.), *Proceedings of the eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from <http://www.stat.auckland.ac.nz/~iase/publications.php>
- Fidler, F. (2006). Should psychology abandon p values and teach CIs instead? Evidence-based reforms in statistics education. In A. Rossman & B. Chance (Eds.), *Proceedings of the seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from www.stat.auckland.ac.nz/~iase/publications.php?show=17
- Gould, R., Davis, G., Patel, R., & Esfandiari, M. (2010). Enhancing conceptual understanding with data driven labs. In C. Reading (Ed.), *Proceedings of the eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from <http://www.stat.auckland.ac.nz/~iase/publications.php>
- Hesterberg, T. (2006). Bootstrapping students’ understanding of statistical concepts. In G. Burrill (Ed.), *Thinking and reasoning with data and chance. Sixty-eighth National Council of Teachers of Mathematics Yearbook* (pp. 391–416). Reston, VA: NCTM.

- Hesterberg, T., Moore, D., Monaghan, S., Clipson, A., & Epstein, R. (2009). Bootstrap methods and permutation tests. In D. Moore, G. McCabe, & B. Craig (Eds.), *Introduction to the practice of statistics* (6th ed., pp. 16-1–16-60). New York, NY: Freeman.
- Hjalmarson, M., & Lesh, R. (2008). Engineering and design research: Intersections for education research and design. In A. Kelly, R. Lesh, & K. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 96–110). New York: Routledge.
- Holcomb, J., Chance, B., Rossman, A., Tietjen, E., & Cobb, G. (2010). Introducing concepts of statistical inference via randomization tests. In C. Reading (Ed.), *Proceedings of the eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from <http://www.stat.auckland.ac.nz/~iase/publications.php>
- Jones, P., Lipson, K., & Phillips, B. (1994). A role for computer intensive methods in introducing statistical inference. In L. Brunelli & G. Cicchitelli (Eds.), *Proceedings of the First Scientific Meeting of the International Association for Statistical Education* (pp. 199–211). Perugia, Italy: University of Perugia.
- 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: Springer.
- Simon, J., Atkinson, D., & Shevokas, C. (1976). Probability and statistics: experimental results of a radically different teaching method. *American Mathematical Monthly*, 83(9), 733–739.
- Sotos, A., Vanhoof, S., Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2, 98–113.
- Tintle, N., VandenStoep, J., Holmes, V., Quisenberry, B., & Swanson, H. (2011). Development and assessment of a preliminary randomization-based introductory statistics curriculum. *Journal of Statistics Education*, 19(1). Retrieved from www.amstat.org/publications/jse/v19n1/tintle.pdf
- Wood, M. (2005). The role of simulation approaches in statistics. *Journal of Statistics Education*, 13(3), 1–11. Retrieved from www.amstat.org/publications/jse/v13n3/wood.html

Chapter 25

Undergraduate Students' Conceptions of Variability in a Dynamic Computer-Based Environment

George Ekol and Nathalie Sinclair

25.1 Introduction

The goal of this study is to shed light on how undergraduate introductory statistics students connect their everyday experiences to the meaning of statistical variability in static and in dynamic mathematics environments. We chose variability for two main reasons. First, from a teaching and learning perspective, variability can be seen as a gateway to understanding other concepts in statistics (see e.g. Garfield & Ben-Zvi, 2008; Rossman, 1996). For instance, in descriptive statistics, reasoning about graphs, mean, median, and standard deviation clearly requires an understanding of variability. Moreover, concepts which are usually taught in inferential statistics such as confidence intervals, hypothesis tests, sampling distributions, regression analysis, and p -values also necessitate a good appreciation of statistical variability. Second, we are interested in developing and using appropriate computer technologies for building concepts in mathematics and statistics.

A number of studies have already drawn attention to the need to help students take a balanced approach towards learning statistics (see e.g. Fischbein, 1987; Makar & Confrey, 2005). By balanced approach, we mean having students build meaning of the concepts they learn, along with the standard procedures. A conceptual framework we have called individual meaning building (IMB) for analysing our data is developed in the next section. The main idea in the IMB framework comes from Fischbein's (1987) original idea on intuition building. Intuition building is further

G. Ekol (✉) • N. Sinclair
Simon Fraser University, Burnaby, BC, Canada
e-mail: gle1@sfu.ca; nathsinc@sfu.ca

developed and used by Makar and Confrey (2005) in their study with pre-service teachers. As mentioned earlier, our goal is to gather information on undergraduate students' multimodal communication about the concept of variability, in both static and dynamic mathematics environments. In the next section, the IMB conceptual framework is developed together with a summary of related studies.

25.2 Framework and Related Studies

Several studies have been conducted on the topic of variability at different school levels (e.g. del Mas & Liu, 2005; Garfield & Ben-Zvi, 2008; Reading & Shaughnessy, 2004). Overall, studies at the undergraduate introductory statistics level seem to indicate that students are more likely to describe measures such as mean and standard deviation using symbolic representations than saying these in their own words to explain variability in a data set (del Mas & Liu, 2005; Garfield & Ben-Zvi, 2008). In their study, del Mas and Liu (2005) designed a computer program for studying variation in data sets. The students used the software to coordinate how the mean varied with standard deviation in different pairs of distributions. The researchers report that although in general, students moved from a process-based notion of standard deviation, many of them did not use the dynamic imagery of the mean and standard deviation to reason about variability. Instead, they resorted to single values of the mean, rather than reasoning about the size of standard deviation.

Makar and Confrey (2005) point to a number of studies that have advocated for learning environments which support students to develop their own intuition and meaning about concepts that they learn. The researchers suggest that while formal procedures are necessary for developing efficient means to problem solving at higher levels of mathematics, the procedures should be carefully taught so that students are conversant with the underlying concepts. Makar and Confrey (2005) refer to Fischbein's (1987) work on intuition. According to Fischbein (1987), intuition building requires personal experience and involvement in practical or theoretical activities. He further submits that developing intuition implies situations in which a student is asked to evaluate, conjecture, devise, predict, and to check solutions. Another important aspect of Fischbein's theory is the role of visualization in intuition building. He argues that what one cannot imagine visually is hard to realize mentally:

it is a trivial affirmation that one tends naturally to think in terms of visual images and that what one cannot imagine visually is difficult to realize mentally. [...] visualization embedded in an adequate cognitive activity remains an essential factor contributing to an intuitive understanding (p. 103).

More recently, researchers have drawn attention to the importance of dynamic visual imagery in mathematical thinking and learning (see Presmeg, 1986; Sinclair & Schiralli, 2003; Sinclair & Gol Tabaghi, 2010). This attention to the dynamic visual images has been inspired in part by studies of learning environments involving dynamic geometry environments and by the growing attention to the role the body

plays in mathematical meaning-making (see Arzarello, Paolo, Robutti, & Sabena, 2009). Building on Makar and Confrey's (2005) argument that "emphasis must be on *building meaning*, not [...] assuming that standard procedures or terms can themselves carry the meanings of the underlying concepts" (p. 31), we propose a framework of IMB that seeks to articulate and develop the personal, intuitive meanings that students develop in working with statistical concepts. This involves attention to the metaphors and nonstandard language that learners use. For example, Makar and Confrey (2005) found that pre-service teachers described data in terms of triads (low-middle-high), modal clumps (points where majority of data were grouped), and distribution chunks (unique units of data distributed within the overall distribution). In our study, for example, descriptions of the mean may include dynamic visual imagery of a self-adjusting fulcrum or seesaw. For standard deviation, descriptions may focus on the distances of individual data points from the mean. IMB also involves attention to learners' non-verbal forms of communication, such as gestures, which have been shown to play an important role in the students' constructions of mathematical meanings (e.g. Radford, 2009).

25.3 Design and Methodology

Our study was carried out at a university in North Western Canada. We interviewed a total of 15 undergraduate students, 5 male and 10 female, all registered in an introductory statistics course. Selecting participants in the study was non-random, but we tried to be as representative as possible by selecting students across different majors: Actuarial Science, Business, Engineering, and Health Science. Participation was voluntary and not linked to final grades. Before participating in the interviews, the students had already covered introductory statistics concepts related to our study, as ascertained by the first author who worked as a tutor in the statistics lab. We interviewed each student one 30–45 min. Each interview had three parts such that in the first part, students were asked to describe their understanding of the mean and standard deviation; in the second part, the students were invited to interact with a dynamic geometry sketch designed by the authors, and to answer questions related to the sketch; in the third part, the students were again asked to describe their understanding of the mean and standard deviation. In this chapter, we focus our analysis on the second part of the interview because we are more interested in how students construct meanings of variability from their physical and visual interactions with the dynamic sketch. We analyse four cases, which we believe represent the diversity of responses in our data collection.

25.3.1 Design of the Dynamic Sketch

The sketch consists of a number line along which six data points are positioned, as in Fig. 25.1a. We used only six data points in order to create a simple model which could still evoke different kinds of distribution. On each data point, a square is

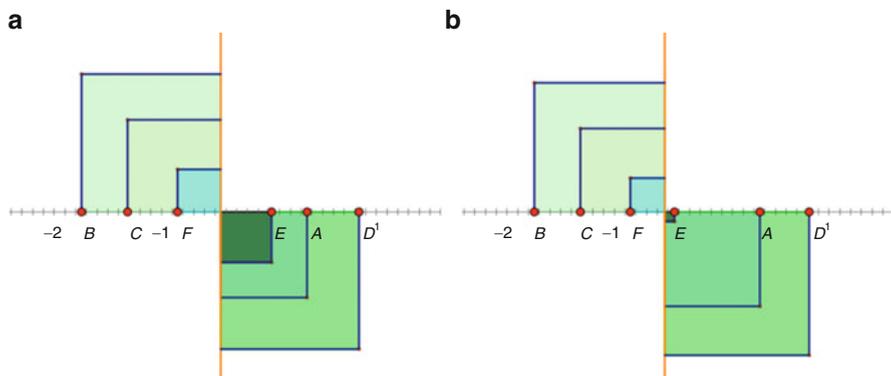


Fig. 25.1 (a, b) The dynamic model of mean and standard deviation

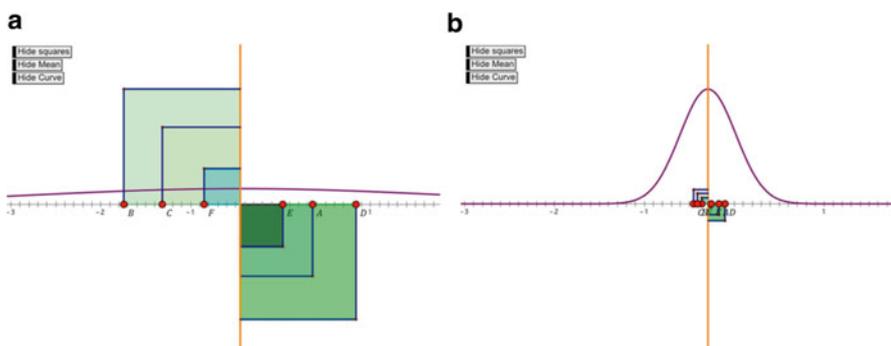


Fig. 25.2 (a, b) The Gaussian curve changes peak as data vary on horizontal axis

constructed that represents the magnitude of the standard deviation with respect to the mean, which is represented by the vertical line. As a data point is dragged along the horizontal axis, the square changes in size as does the location of the mean, as seen in Fig. 25.1b.

The dynamic model also includes an option to show the Gaussian curve for the six data points as shown in Fig. 25.2a. When the data points are dragged, it is possible to create something approaching a normal curve, that is, a Gaussian with mean of 0 and standard deviation of 1.

25.3.2 Interview Protocol

In the second part of the interview, students were shown the sketch and asked to predict what would happen to the mean and standard deviation as the six data points were dragged on the horizontal axis. We expected students to predict that as they

dragged the data points closer together, the value of the mean would decrease as well as the standard deviation (the size of the squares on each data point). But if the data points were dragged farther away from each other, the mean and standard deviation would both increase on the right side of the mean line (vertical line); but on the left side of the mean line, the mean value decreases as standard deviation increases. The students were then asked to drag the data points and check their predictions.

In the second task, the Gaussian curve was shown and students were asked to predict how it would change as the data points were dragged along the horizontal axis. We expected students to predict that as the data points were dragged away from each other, the curve would get flatter; but as the points were dragged closer together, the curve would rise. Students were then invited to drag the points and check their predictions. The second task proved more difficult for the students than the first task in the sense that students had noticed in the first task, that when they dragged all data points together the squares seemed to converge to a single point, which behaviour they also seemed to attribute to changes in the normal curve.

The interviewer never used the word variability. However, the goal of the tasks was to find out *how* students would describe the changing positions of the data points as well as the changing values of the mean and the squares representing the standard deviation. More specifically, we were interested in whether the students would talk about the variability of the data in terms of their relative positions (how close together or far apart) and in terms of their distances from the mean.

25.4 Results

As explained earlier, we asked participants to talk about the mean and standard deviation before introducing them to the dynamic models. We present data for four participants: Kimberly, Remy, Bonita, and Yuro (not real names, as required by ethics).

25.4.1 *Mean and Standard Deviation: Part I of Interview*

After a brief introduction, the interviewer asked students to talk about the *mean* and the *standard deviation*.

Kimberly: the mean is the answer to a formula where we add up [...]...so mean is like a specific number [...] mean is more specific, it's calculated, [...], makes me think like when we weigh something and the mean is like the exact center, the exact middle [...]. I see standard deviation in graphs [...] one, two, three, negative one, negative two [...]. You can calculate standard deviation.¹

¹ In the transcripts we use "[...]" to indicate that some parts of the sentence is not included to make it shorter, while "..." is used for pauses. [Words in italics inside the square brackets are added by the interviewer].

- Remy: the mean [...] the statistical thing [...] *[it involves]* pretty much adding all the numbers and dividing by the number of numbers. *[On standard deviation]* I would think of a symbol, yeah, [...] but this is the variance...yeah I would think like that, there is lots of formula, more abstract concept, that's what comes to mind.
- Bonita: *[About the mean]* I immediately think of the averages because simple definition of the mean I guess would be adding up all the numbers in the data set [...]. *[On standard deviation, Bonita referred to what she said before about the mean]* I think that is kind of related to the mean. As I said before, if you can figure out the mean of a data set [...] then you can derive the standard deviation [...].
- Yuro: *[On the mean Yuro's response had only two words]* Mean? average *[he laughed. Similarly, Yuro did not talk much about standard deviation]* ... spread, like same things, deviation and standard deviation are like same things to me.

The verb “add” is used by three of the four students with respect to the mean, indicating that the students were thinking of the mean in terms of a computational activity that generates a single number. This thinking does not take into account the notion of variability. Yuro's response (“average”) shows less emphasis on the process type of thinking, but it is not clear whether his sense of average relates to variability.

In terms of standard deviation, the participants' answers are less precise. Kimberly's reference to graphs and distribution of points on a number line does express a sense of the variability of data in terms of standard deviation. But she quickly returns to computational-oriented type of thinking, saying “you can calculate standard deviation”. Bonita also mentions that standard deviation can be derived, but emphasizes its relationship to the mean. Remy mentions the notion of “variance” but connects standard deviation to “a symbol” and “lots of formulas”. Yuro mentions the word “spread” in talking about standard deviation, which can be seen as a way of describing the data set in terms of their relative positions, and perhaps alluding to the concept of data variability.

25.4.2 Mean and Standard Deviation in Dynamic Mathematics Environments

We present the interview data in the same sequence as we did with the static data, and analyse each in terms of the emergence of the participant's idea of variability.

- Kimberly: Well, I guess when I move the points to the right the squares will increase *[Kimberly actually pointed to the left side with her right index finger at the very moment she said the squares will increase]*.
- Int: Which is your right?

- Kimberly: Oh no to the left, and when I move it to the left the square area will decrease
- Int: Why do you say that?
- Kimberly: Because the farther away the point is from the centre, then the greater area it has.
- Int: [Although Kimberly's last statement was correct, she was not clear about the left or right directions from the centre, so the interviewer asked her to clarify her prediction]. Ok, so your claim is that when you move the data points to the left side of the center line, the area will increase. What if you move the points to the right side?
- Kimberly: When I move them [the data points] to the right, then the area will decrease [she pointed to the right side using her right index finger as she said 'right']
- Int: Ok, now you can test. [As Kimberly dragged the point to the left side, the square increased in size, as she had predicted. The interviewer then said] "You are right ... what does that tell you about data and the center?"
- Kimberly: When data points are [dragged] farther away from the center, then the slope, ah, their values will be greater but when the points are closer to the center, they [their values] tend to be smaller.

Kimberly attended to the changes in the mean and the standard deviation as she dragged the data points along the horizontal axis. As she found confirming evidence of her hypothesis, by dragging the points. She seemed to pay close attention to the variability in the data points as the mean and standard deviation changed.

- Int: How about the curve? What will it look like as you drag the points?
- Kimberly: [...] I guess as I move [data] points, the line [the curve] will also rise [Fig. 25.3 a shows hand movement when she said 'rise']. So when standard deviation increases the curve [the peak] will also increase.

Kimberly did not specify how the curve will rise—was it by moving the data points away from the centre or moving them toward the centre?—but she did relate

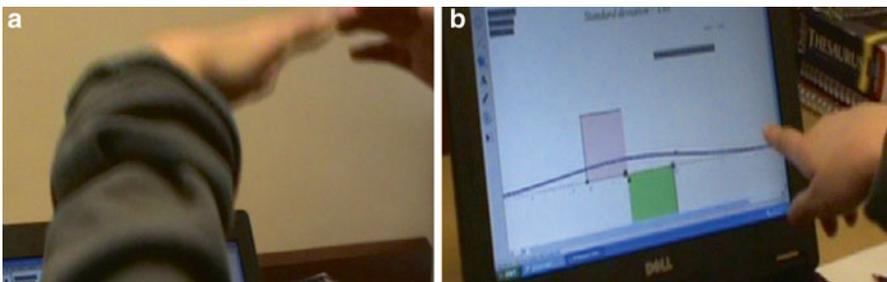


Fig. 25.3 (a) The curve will rise. (b) It's like a hill

the changing values of the points to a change in the curve. Her last statement about standard deviation was the opposite of our expectation, as Kimberly herself confirmed.

Int: Ok, now go ahead and check.

Kimberly: Opposite (she laughed), opposite, oh opposite to what I, to what I got. It's like a hill! [Fig. 25.3b, she moved her index finger in the air from left to right to show a rising curve].

We were not surprised that Kimberly did not use her earlier prediction about the squares to talk about the curve. As we noted earlier on, connecting the Gaussian curve with standard deviation and variability in the data set was challenging to the students. But after checking her prediction by dragging the data points closer to the centre, Kimberly was able to confirm that as the mean and the standard deviation decreased, the Gaussian curve began to rise.

Remy's interaction.

Int: [*Remy did not make a prediction about the mean and standard deviation as data were dragged on the horizontal axis. Instead, he asked to try out the model. After taking more than five minutes exploring the dynamic model, he remarked*]

Remy: This is kind of, this stuff is interesting.

Int: What do you mean?

Remy: It's not something you see everyday, I mean, like... because you have the formula [...] but they don't show it, they don't show it like this. I mean this is like, what is going on here, I mean this is very visual right, so it's easy to remember actually. Say oh yeah, the squares are moving right now, what are they doing? Then you say ok, if I move this one over here, then and so on. Um, so it is easy to remember when you see it like this actually.

Remy was able to recognize the use of the sketch for showing variability in the data points, by observing how the squares were varying with respect to their distances from the mean. The changing values of the data points and the corresponding mean and standard deviation seemed to provoke Remy to question, to check, and to confirm variability in the data set.

Bonita's interaction.

Int: [*Bonita predicted that the squares would get bigger as data points varied from the center, but the interviewer wanted her to talk of variability in terms of the mean and standard deviation*].

Bonita: Well, if I take one [data] point [...] and move it away from the center [to the left side of the center], [...] I think the square will move this way [to the left side], um ... so it will get a bit bigger ... and if I move it toward [the center], it will become a bit smaller because I'm going more toward the center and I'm going in the positive direction, that's what I think.

Bonita: If the mean gets smaller, the standard deviation gets larger.

- Int: If the mean gets smaller, the standard deviation gets ...?
- Bonita: It would get larger, but by not a significant amount, so it will be very little with respect to that.
- Int: Now you can test them.
- Bonita: Oh! So, both of them got larger, yeah ok, I thought the mean would get smaller and the standard deviation would get larger, but actually both of them are increasing.
- Int: [About the curve described in Fig. 25.2a, b, Bonita made the following predictions].
- Bonita: I think if you move the point [...] here, toward the mean here [*she pointed toward the center with her right index finger*], it will make the shape rise a bit more.
- Int: Why do you say that, what makes you say that?
- Bonita: Because already at the D point over here (Fig. 25.4a) [...] this part [...] nearby the D point is really flat.
- Int: But it could well rise as you move the data points away [...]. Ok, go ahead and test. [*As she dragged the points closer to the center, the curve began to rise, as shown in Fig. 25.4b*].
- Bonita: I didn't know that. Well, when I was working on the other example [*with the squares*], it also had six data points and they were clustered... and [*now*] I see it more of a normal distribution [*she drew Gaussian curve in space with her right index finger as shown in Fig. 25.4b*].

After interacting with the dynamic models, Bonita gained awareness about how standard deviation changed with respect to the mean. Her statement, "Oh! I didn't know that", showed that she was seeing something new. It is interesting that Bonita described the curve as the normal curve, not say, using the metaphor of a hill like Kimberly did. The transition from the flat Gaussian curve in Fig. 25.4a to the normal curve Fig. 25.4b shows a reduction in variability of the data points. Bonita was able to describe the shape of the data set as a whole in terms of the relative position of each data point to the other.

Yuro's interaction.

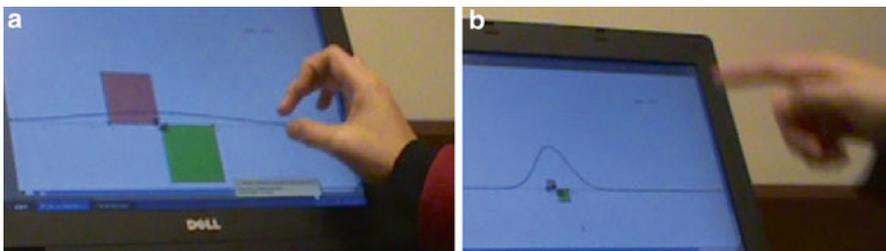


Fig. 25.4 (a) This part is really flat. (b) I see more of a normal curve

Int: [Yuro started with prediction on how the mean and standard deviation would vary as data points were dragged on the horizontal axis.]

Yuro: The squares, if you start moving it this way [moved his right index finger from left side toward center of the sketch, as in Fig. 25.5a], the squares would decrease.

Int: Why is that?

Yuro: Oh, because over here you have [...] like a central point (Fig. 25.5b). And then these are distributions around the central point, and if you are going to move this way [he indicated the left side], your distribution is going to increase. [Yuro later dragged a point to the left side and discovered that the squares did not move as he predicted]

Yuro: So I take this and go ... Oh, ok!

Int: Why did you say oh?

Yuro: Coz I was wrong (he laughed) [...]. Because I only thought that this [horizontal length as shown in Fig. 25.5a] was gonna increase but not this one [the vertical length of the square], but now that I think about it makes sense.

Yuro predicted a decrease in the size of the squares as the data points were dragged away from the centre on left side. After checking his prediction on the dynamic model, Yuro seemed to connect well how the size of standard deviation and the mean explained the variability of the data points. His comment, “and now that I think about it makes sense”, provides some evidence to his awareness of the relationships.

25.5 Summary

We began the study with a hypothesis that dynamic mathematics environments can help promote a more physical and temporal understanding of statistical variability. Using IMB conceptual framework to analyse the video data, we were able to show that, when interacting with the sketch, the students spoke about the mean and the standard deviation in terms of thier physical and numerical changes on the sketch,

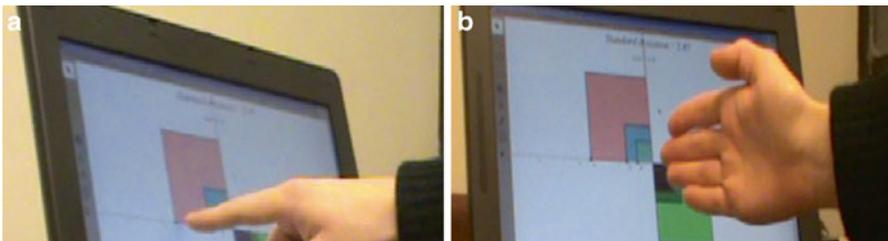


Fig. 25.5 (a) This way, the squares increase. (b) Here you have a central point

e.g., “I thought the mean would get smaller and standard deviation would get larger, but actually both of them are increasing” (Anita). The changing values of the data points, along with the associated changes in the size of the squares and the position of the mean seemed to invite the students to attend to the relationship between the mean and the standard deviation. In addition, by dragging the data points, they seemed to attend to the distance between the data points and the mean as well as to ways in which the data points were positioned one relative to the other. In describing their observations, the students drew on dynamic language (e.g., moving, increasing, decreasing, rising, etc.), on metaphors (hill, cluster), and on gestures. These actions shifted students' thinking away from using a formula to calculate the mean or standard deviation that was prevalent at the beginning, to thinking about the mean and standard deviation more qualitatively in terms of variability.

25.5.1 Limitations

We started with 15 students and reported only on 4 cases. We recognize that this imposes some restrictions to how generalizable our claims can be.

References

- Arzarello, F., Paolo, D., Robutti, O., & Sabena, C. (2009). Gestures are semiotic resources in the mathematics classroom. *Educational Studies in Mathematics*, 70(3), 97–109.
- del Mas, R., & Liu, R. J. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4(1), 55–82. Retrieved from www.stat.auckland.ac.nz/serj/.
- Fischbein, E. (1987). *Intuition in mathematics and science: An educational approach*. Dordrecht, The Netherlands: D. Reidel.
- Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.
- Makar, K., & Confrey, J. (2005). “Variation-talk”: Articulating meaning in statistics. *Statistics Education Research Journal*, 4(1), 27–54 [Online: www.stat.auckland.ac.nz/serj/].
- Presmeg, N. C. (1986). Visualization in high school mathematics. *For the Learning of Mathematics*, 6(3), 42–46.
- Radford, L. (2009). Why do gestures matter? Sensuous cognition and the palpability of mathematical meanings. *Educational Studies in Mathematics*, 70(3), 111–126.
- Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201–226). Dordrecht, The Netherlands: Kluwer Academic.
- Rossmann, A. J. (1996). *Workshop statistics: Discovery with data*. New York: Springer.
- Sinclair, N., & Schiralli, M. (2003). A constructive response to ‘Where mathematics comes from’. *Educational Studies in Mathematics*, 52(1), 79–91.
- Sinclair, N., & Gol Tabaghi, S. (2010). Drawing space: Mathematicians' kinetic conceptions of eigenvectors. *Education Studies in Mathematics*, 74(3), 223–240.

Chapter 26

Transforming Statistics Education Through ICT Application

Tae Rim Lee

26.1 Introduction

Korea National Open University has been utilizing one-way education delivery systems during its history of distance education since 1972. In this one-way mode of the systems, isolation of students in their learning process has been the most important problem to solve. ICT application such as e-Learning or m-Learning is an alternative instructional model that enables students to have more interaction with their instructors and peers by providing more accessibility to multimedia learning resources than the conventional delivery system provides.

Since 1997, KNOU have been carrying out e-learning projects as a member of KVC (Korea Virtual Campus) consortium, which consists of 10 ordinary universities and ITCU (Information Technology Cyber University) consortium, which consists of 36 universities in Korea (KVC, 2003). These consortium projects have been mainly carried out for small classes having 70 students or fewer, while using the start-up e-Learning Management System (LMS).

On behalf of starting online graduate school programs of four departments, e-Learning Center was established with 24 members including educational technologists, web programmers, web designers, computer system analysts in 2001. During the next year 2002, the e-Learning hub site, “e-Campus,” was launched, and 38 e-Learning courses were developed, funded by Ministry of Education (MOE) & Human Resources Development. During 2004 to 2005, KNOU e-Learning Center developed eight international e-Learning courses in English, funded by MOE:

T.R. Lee (✉)

Department of Information Statistics, Korea National Open University, Seoul, South Korea
e-mail: trlee@knou.ac.kr

courses on Korean History, Korean Culture & Art, Economic Development and Economic Policy in Korea, Click Korean, Statistics, Water Quality Test Method, Introduction to Computers, History of Economy (<http://elic.knou.ac.kr/>).

The e-Learning system for distance education has improved the lack of two-way communication and repeatability of learning, the main weaknesses of the conventional media such as TV, radio, and written text (Jung & Kim, 2002). The e-Learning system has extended the opportunity of learners by operating a variety of curriculums on the basis of e-learning (Clark & Mayer, 2003). The 2005 project was to evaluate effectiveness of the e-LGD project with class diversity that was launched for undergraduate students in the first semester in 2004 and to make suggestions for its future expansion to all regular courses. The volunteer students of each course had a chance to access e-learning contents and relevant learning materials, and were also given some announcements and chances to interact with their professors and colleagues during the one semester. The survey was included in e-LGD project for the course evaluation to identify the current status of e-Learning and the improvements to be made for more effective e-Learning. The findings in this study surveyed by faculty members and students were analyzed in terms of learning contents, course management, and administrative support.

In December 2008, KNOU launched the mobile learning system under the MOU with a major Korean telecommunication company, KT. The mobile learning and ubiquitous learning systems for distance education enable any aspiring students to study anywhere, anytime with the Internet and multimedia systems using portable electronic devices. m-Learning can become an ordinary part of open and distance learning (ODL) for lifelong education and distance learning in the near future. Several projects were launched to evaluate the effectiveness of e-learning courses and to suggest future improvements of e-Learning courses and future views of a more advanced education system of mobile and ubiquitous learning systems. In the future knowledge-based society with a rapid change of educational circumstances and paradigm, distance education using ICT technology can satisfy the educational needs in various levels of learners. KNOU has provided students with distance statistics education contents through broadcasting and ICT-adopted media through Internet.

Mobile technologies, including mobile devices and wireless Internet services, have the potential to introduce new innovations to education with m-learning, a new form of education using the mobile Internet system and handheld devices. This can offer students and teachers the opportunity to interact frequently with and gain access to educational materials independently of time and space. The 2009 study made some considerable suggestions for preparing for the future of distance education based on mobile and one-step-further advanced ubiquitous learning systems.

Internationally, KNOU was assigned as the coordinator of e-ASEM network under the research theme, "ICT Skill, e-Learning and the Culture of e-Learning in Lifelong Learning," among the four education and research network themes of ASEM LLL (Life-Long Learning) in May 2005. The project team, therefore, plans

to establish an online community for sharing ICT skills and e-learning-related educational knowledge and researches among the ASEM LLL member countries (<http://asem.knou.ac.kr/>).

In 2006, KNOU organized the Asia-Europe Colloquy on University Co-operation on “e-Learning for Higher Education” with the theme, “Challenges and Opportunities,” where 87 delegates from Belgium, Brunei, Cambodia, China, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Indonesia, Japan, Laos, Latvia, Lithuania, Malaysia, Malta, Netherland, Philippines, Poland, Singapore, Slovakia, Thailand, UK, and Vietnam attended and had a highly productive and successful colloquy (www.elearningcolloquy.org).

In 2007, KNOU organized an e-ASEM follow-up meeting at Seoul, Korea, where we exchanged our detailed experiences and knowledge of ICT application to ODL and LLL (<http://infostat.knou.ac.kr/eASEMnetwork2007/>).

26.1.1 Overview of Situation

Since its opening in 1972, KNOU has been growing as the only one mega-university in Korea for ODL with a considerable scale for the past 30 years. KNOU consists of 4 colleges including 22 departments. It has approximately 183,400 students and has turned out 290,000 graduates so far. Also, it has opened a graduate school based on e-learning with 6 departments and 568 students. The large number of students reveals the high and dynamic demand for lifelong learning of the Korean society. KNOU has managed various curriculums corresponding to such a high demand for lifelong learning. However, recent sociocultural and environmental changes related to the ODL provide many suggestions for the new direction for the development of KNOU.

First, the major delivery system of statistics education has changed as information and communication technology develops. KNOU has been using one-way delivery systems such as TV, radio, and audio cassette tape. However, the developments in computer science and communication technologies opened the path to a two-way delivery system that enables learners to actively participate in their learning process (Kolondner & Guzdial, 1996).

Second, it was a hot issue that several cyber universities conferring a bachelor’s degree. Since 2001 in Korea, 16 cyber universities have been established (Ministry of Education and Human Resources Development, 2003). As a result, the variety of lifelong education institutes brought about competition among the conventional distance education institute of KNOU and those other cyber universities.

Third, there has been an increasing tendency in the variety of the students in KNOU. In the past, most of the students of KNOU were those who had no chance to enter a university after graduating from high school, but recently the proportion of those who enter KNOU for reeducation or transition into a different major after a bachelor’s degree has grown considerably as in Table 26.1 and Fig. 26.1.

Table 26.1 Distribution of bachelor students' admission to KNOU

Year	2005	2006	2007	2008	2009
Number bachelor admission	20,835 (19.9 %)	16,699 (6.8 %)	16,238 (15.9 %)	15,671 (14.3 %)	9016 (8.3 %)
Total	104,724	99,061	102,268	109,311	108,367

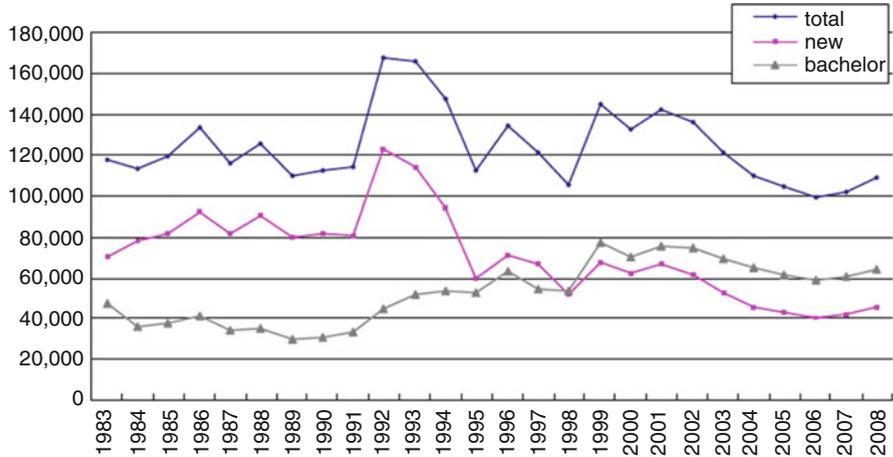


Fig. 26.1 Trend in students' enrollments in KNOU according to their education level

This implies that the needs for a flexible teaching–learning system corresponding to the varying levels of the students should be analyzed.

KNOU decided to launch its own full-scale e-learning project to overcome this challenging problem. KNOU has been running its own e-Learning courses and adopting the cross registration system through KVC and ITCU, the consortiums consisting of 10 universities and 36 universities, respectively.

Additionally, the number of courses developed for e-Learning among the entire undergraduate curriculums reached up to 153 (18.9 %) by developing courseware and supplementary learning resources. The number of students in e-Learning classes of KVC and ITCU ranges only from 60 to 70, and thus it is inappropriate to directly apply the course management strategy designed for a class of this size to the ordinary class of KNOU having over 10,000 students in each class.

Therefore, KNOU recognized the need for a study on an e-Learning course management plan considering various class sizes. The objective of the 2005 project was to develop a future plan for applying the e-Learning model to regular courses of KNOU after evaluating the e-Learning contents that were run in ten courses for the college students during the first semester in 2004. To accomplish this objective, in the project, we analyzed the e-Learning contents, the course management and student support, and the present situations of institutional and administrative support, and then produced an improved vision and practical plans for the future.

26.1.2 Earlier e-Learning Projects in KNOU

26.1.2.1 Objectives of e-Learning Adaptation

The objectives of e-learning adaptation in KNOU courses are as follows:

First, to provide students with easy accessibility to learning resources. One of the typical advantages of distance education is flexibility in learning as to time and space, which means that a student can learn anywhere and anytime. This provides usefulness to those who used to have difficulty in following the fixed regular course schedule (KNOU, 2003).

Second, to provide various learner-oriented materials comprehensively. According to the research results, the learners could access fruitful self-study materials and study them comprehensively by e-Learning (KNOU, 2003). It can be said that the management and delivery of study materials through ICT is an effective way to meet the learner's needs for interactivity.

Third, to motivate the students to become a self-regulated learner. The self-regulated learning ability can be an important factor greatly affecting whether e-learning could be successful or not since the learner takes the initiative in making a decision about the study process and method. The self-regulated learning has the characteristics of meta-cognitive strategies that further, manage, control, and improve one's learning through setting the goals of study, reviewing, evaluating, and managing oneself (Knowles, 1975). The self-regulated learning is the ability to include a motivating element to continue one's learning and a behavioral element to practice (Zimmermann, 1990). Since e-Learning requires learners to play an active role in their learning process, they naturally develop self-regulating ability.

26.1.2.2 Readiness for e-Learn

According to the data from the National Computerization Agency (2004), in 2002, the number of Korean population who own a personal computer reached 49 per 100, and the rate of the population using the Internet 61 %, and that of the high-speed network 23.3 %. Our international information index holds the eighth rank. Additionally, according to the data from Korean Ministry of Information and Communication (2004), the Internet and high-speed network use rates increased to 74.8 % and 24.2 %, respectively. These data imply the sufficiency of technical infrastructure, which is the ideal condition for the adaptation of e-learning.

To provide the most effective e-Learning service, the following conditions need to be met at the three levels of the participants in the e-Learning system including learners, instructors, and the service organization as follows.

First, for preparation at the learner level, how well learners can prepare for learning is influenced by how much they can use the Internet and the high-speed network. According to the survey for 'A Study on the Actual Condition of the Use of Learning Media of the Students who are attending Korea National Open University

(KNOU, 2004),’ where 102,940 students (52 %) among 196,402 who registered for the first term of 2003 responded, the rate of the students who were using the Internet was 95.3 %, and almost all of the students could have access to academic information and learning information on the Internet. In addition, the high-speed network use rate of those students reached 81.9 % and the LAN use rate 12.8 %. Thus, it was indicated that 93.7 % of the students had no difficulty using the variety of multimedia learning resources. In the 2007 student survey, 83 % of registered students responded that they can take class on the high-speed Internet system, and 84.7 % responded their main communication and information delivery tools were computers. These findings provide the grounds for developing high-quality learning contents and utilizing them actively.

Second, for preparation at the instructor level, how much experience instructors have in e-learning contents development and course management influence to the quality of e-Learning contents and evaluation results of course management. In 2004, 55.9 % of the professors of KNOU had an experience in e-learning contents development. In particular, 90.0 % of the professors of the Faculty of Science had contents development experience. These statistical findings can be interpreted as the possibility of e-learning contents utilization initiated by the professors with enough experience.

Third, the organization level preparation is one of important factors. By 2004, KNOU developed e-learning contents for 103 courses out of total 554 courses annually opened by the faculty, which amounted up to 18.9 %, and had e-learning staff, who were wholly responsible for e-learning contents development, course management, consultation, faculty training, and educational program management, by establishing the e-Learning Center at 2001 to build an effective e-learning support system.

26.1.2.3 Major Issues

The variety of educational demands and the change of paradigm in statistics education in ODL were strong motivations to renovate the educational media by ICT application. As the students’ access to the Internet and use of ICT has been increasing rapidly, educational space in this area is expected to be enlarged significantly and we need to explore various levels in the teaching–learning system. Standardization and quality control process will be needed to support development of high-quality e-contents and m-contents.

- Variety of educational demands
- Change of paradigm in distance education
- Increased access to the World Wide Web
- Enlargement of educational space
- Explore various levels in the teaching–learning system

- Standardization of e-contents and m-contents for quality improvement
- Necessity for complete transfer from e-contents to m-Learning contents

26.1.3 KNOU Ubiquitous Learning Campus

KNOU ubiquitous learning campus was launched in December 2008. Mobile technologies using mobile devices and wireless Internet services have the potential to introduce new innovations in the area of statistical education, a new form of statistics education using the mobile internet system and handheld devices, which can offer students and teachers the opportunity to interact with and gain access to educational materials independently of time and space. This study made some considerable suggestions for preparing the future of open and distance education based on mobile technology and one-step-further advanced ubiquitous learning in statistics education. Figure 26.2 shows the title window for KNOU mobile learning cooperated with Korean telephone company KT under the support system for student's mobile learning like in Figure 26.3.

Mobile Campus in Hand

Using a new information sharing device, the mobile phone, makes renovation to U-Campus, and the technology solutions promote the renovation to the new paradigm of KNOU U-CAMPUS in hand. It provides composite solutions of on and

Fig. 26.2 Title window of U KNOU m-Learning Campus



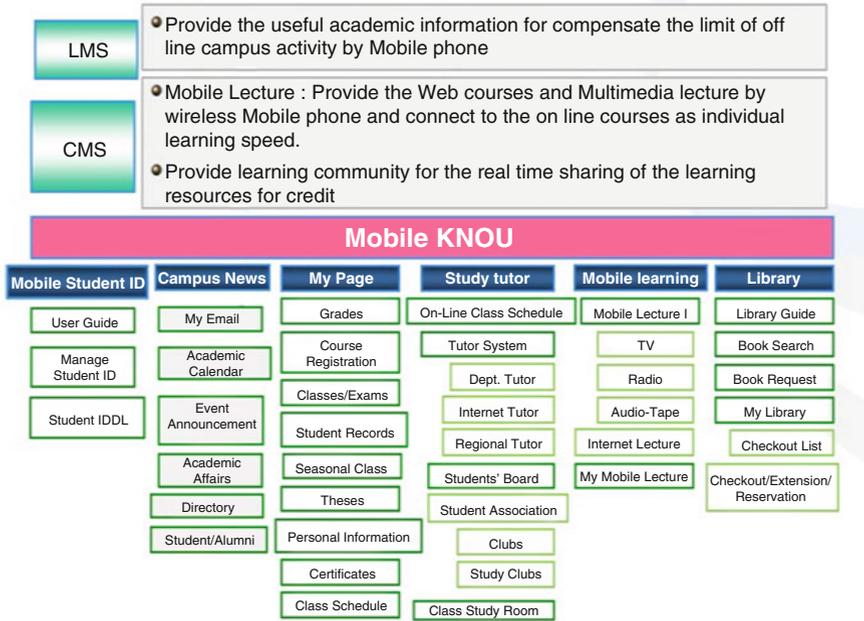


Fig. 26.3 Support system for the student’s mobile learning

off line direct connections between the LMS for KNOU U-Campus and KT Mobile Solution.

For the high-quality m-Learning, it should be continued to evaluate and give feedback to the ODL statistics learning resources under the team approach which brings educational technologists, computer analysts, web programmers, web designers, and contents specialists together.

26.2 Conclusion

This study intended to draw up plans for introduction of e-learning m-learning courses to the whole undergraduate statistics curriculums in the near future through the evaluation of learning contents, course management and student support, and institutional and administrative aspects.

This study led to the following conclusions and suggestions:

First, there is no meaningful difference in the students’ level of satisfaction with e-Learning contents and e-Learning course management according to class size. Therefore, further studies should be conducted on e-Learning course models according to the various class sizes of KNOU. The e-learning courses should be developed considering various elements such as type of study, class size, and study goal, and standardized management programs according to each model. Also, it is necessary

to conduct further research over how methods of teaching and learning should be implemented according to various operation models of e-Learning courses. Recursive studies are required to supplement the e-learning operation plans by the model through implementing the results of these studies into the actual teaching environment and verifying their efficiency.

Second, improvement in self-evaluation methods is needed for active learning participation by learners. Based on learners' questionnaire survey, the preference is that they tend to have a lesson in a rather passive mode held by the teacher. These results are derived from the evaluation method that evaluates the understanding of lesson contents. However, reflecting on the fact that learning should be reinterpreted through the experience of learners and they are to be able to put their learning in practice, a self-evaluation method that can require a more active participation by learners is needed here.

Third, the incorporated policies for various media such as TV, radio, and e-learning are required. The e-learning can utilize previously developed broadcasting media usefully. Therefore, the broadcast media should be developed as a component consisting of e-learning contents considering that they can be reused for e-learning from the planning stage of TV or radio program development.

Fourth, a systematic study support system should be built. KNOU has been mostly focusing on support for professors, but the actual situation is that the construction of the student support system is not sufficient. Consequently, the type of help and support that learners need should be broken down through in-depth follow-up studies, and the appropriate countermeasures should be groped for. The construction of learner-centered service is the subject that KNOU should concentrate on for the future.

Fifth, a learner-tendency analysis program is required. To develop this kind of program, the information on learners such as their preference and level should be systematically managed and it should be actively applied to the course development and management (Joung & Kwak, 2004).

Sixth, quality should be continuously controlled during the whole process of e-learning course development (Joung & Jang, 2004). Presently, KNOU is controlling quality from the e-learning course development stage through an instructional system design, but it needs to establish a circulative quality control system by confirming whether evaluation results have improved in the next courses.

References

- Clark, R. C., & Mayer, R. E. (2003). *e-Learning and the science of instruction*. Hoboken, NJ: Wiley.
- Joung, Y. R., & Jang, E. J. (2004). A study of quality criteria for excellence of e-learning course. *The Journal of Educational Information and Media*, 19(2), 159-192.
- Joung, Y. R., & Kim, S. Y. (2002). Design template framework of learning object for reusable e-learning contents. *The 16th AAOU Annual Conference Proceeding*.

- Joung, Y. R., & Kwak, D. H. (2004). A study of standardization model of learner information in e-learning. *The journal of Korean Association of Computer Education*, 7(4), 77–91.
- KNOU (2003). The Research on the utilization state of learning media in KNOU.
- KNOU (2004). *A study on the actual condition of the use of learning media of the students who are attending Korea national Open University*, KNOU IDE.
- Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers*. Chicago: Association Press.
- Kolodner, J., & Guzdial, M. (1996). Effects with and of CSCL: Tracking learning in a new paradigm. In T. Koschmann (Ed.), *CSCL: Theory and practice* (pp. 307–320). Hillsdale, NJ: Lawrence Erlbaum Associates.
- KVC. (2003). The 2nd white paper of evaluation for Korea virtual campus. KVC.
- Ministry of Education and Human Resources Development. (2003) 2003 Educational information white book. Korea Education & Research Information Service.
- Ministry of Information and Communication. (2004). The informational statistics of Korea. Retrieved from <http://www.mic.go.kr/index.jsp> [on-line available].
- Zimmerman, B.J. (1980). Self-regulated learning and academic achievement; An overview. *Educational Psychologist*, 25(1), 3–17.

Chapter 27

Selecting Technology to Promote Learning in an Online Introductory Statistics Course

Megan Mocko

27.1 Introduction to the Technology

The purpose of this chapter is to help other instructors who are asked to teach an online course for the first time to learn from the experience of others. At the end of the chapter, several recommendations will be made for instructors who plan on teaching an online statistic course for the first time.

How to deliver content is often the first area of need that instructors must satisfy for an online course. The chapter will discuss SoftChalk, a lesson-building software package that can be used to build lessons with videos, short quizzes, and flash-based activities. In addition to the instructor prepared website lessons, the textbook publisher's website called MyStatLab will also be discussed briefly. Additionally, communication between the student and the teacher as well as communication among students is critical for a good learning environment. Several forms of communication programs will be discussed including programs within Sakai (the course management system) and Elluminate by Blackboard. Elluminate is an online software package that includes video chat, text chat, and an interactive whiteboard. Sakai is an open source course management system that includes a text chat room and discussion board. Conjointly, a brief review of the hardware used in the creation of the recordings for the online course will be covered including the tablet PC and wireless microphone systems. Finally, how do you handle formalized assessment exams in

An extended version of the chapter has been published in: Megan, M. (2013). Selecting technology to promote learning in an online introductory statistics course. *Technology Innovations in Statistics Education*, 7(2). <http://escholarship.org/uc/item/596195sg>.

M. Mocko (✉)

Department of Statistics, University of Florida, Gainesville, FL, USA

e-mail: mmeece@ufl.edu

an online environment? How do you test the student's understanding of the material when you are not physically in the same location? An online test proctoring service, ProctorU, will be discussed in detail.

27.2 Setup

This research was conducted with a class of 67 undergraduates at a large research institute in the United States. The course was taught during the summer over a period of 6 weeks. The stipulation made by the university was that the students would not be required to come to campus for any portion of the class and therefore everything had to be done completely online. There were multiple areas of assessment in the course: exams, daily lesson quizzes, chapter quizzes, small group discussion board assignments, daily homework, and a final project. Each of these assignments utilized technology at some level.

The students were asked, but not required, to complete a 38-question presurvey and a 40-question postsurvey. Out of the 67 students enrolled in the course, 28 started the presurvey and 25 completed it. For the postsurvey, 28 started it and 22 completed the survey. All students who completed the survey were 18 years old or older.

A brief overview of the class demographics from the survey showed that 89.5 % of the students were female and 10.5 % were male and that 47.4 % were sophomores, 42.1 % were juniors, and 10.5 % were seniors in college. The students were also asked if they had taken a statistics class before and 73.7 % had never taken a statistics class while 5.3 % had taken a (non-AP) statistics class and 21.1 % had taken an AP level statistics class in high school. After earning a bachelor's degree, the student's plans for the future included 21 intending to pursue a post-bachelor degree in medical school, veterinary school, graduate school, or law school, while one each intended on entering the work force, military, and Peace Corps. These demographics show that most students that completed the survey were female, had not taken a previous statistics class, and were planning for a post-bachelor degree.

27.3 Background

For more than a decade, statistics educators have been researching how to implement statistics courses online and determining if there is a difference between online courses and traditional courses.

Utts, Sommer, Acredolo, Maher, and Matthews (2003) compared a traditional course to a hybrid course (partially online) that met only once a week. She found that "performance of students in the hybrid offering equalled that of the traditional students, but students in the hybrid were slightly less positive in their subjective evaluation of the course" (2003, p. 1).

Tudor (2006) discussed a course for public health students where she included voice over PowerPoint slides and quizzes for self-assessment. The quizzes were

static quizzes on Word files with answers supplied. She did include discussion board assignments in her course, but concluded that “. . . it appears that the effectiveness of online discussion in a statistics class is still debatable. The biggest factors affecting their success may be the topic of discussion and the quality of the questions” (2006, p. 7).

Everson and Garfield (2008) discussed using discussion boards in her online courses to align her course with GAISE (Aliaga et al., 2005) guidelines. She states:

An important goal of the online course described in this chapter was to align them with the GAISE recommendations. Based on our experiences and observations, structuring our online courses in this manner has resulted not only in student learning but in student satisfaction. (2008, pp. 9–10)

The GAISE guidelines are a series of guidelines for educators teaching statistics in the United States. The GAISE guidelines recommend:

that instructors emphasize statistical literacy and develop statistical thinking, use real data, stress conceptual understanding rather than mere knowledge of procedures, foster active learning in the classroom, use technology for developing conceptual understanding and analyzing data, and use assessments to improve and evaluate student learning.

Mills and Raju (2011) summarized and compared 20 articles about online courses in statistics over the past decade. Mills asserts that:

In the middle to latter part of this decade, more importance was and has been placed on: selecting “appropriate” uses of technology for the online statistics environment, improving interaction among students and the instructor, enhancing the overall learning experience for online students, and conducting formative and summative evaluations to carefully monitor the teaching and learning process. (2011, p. 21)

Additionally, general education literature can also tell us about the important components of an online course. The text, “The Online Teaching Survival Guide: Simple and Practical Pedagogical Tips,” contains a list of ten best practices for online teaching including the following three items; “create a supportive online course community,” “use a variety of large group, small group, and individual work experiences,” and that the instructor should “prepare discussion posts that invite responses, questions, discussions and reflections” (Boettcher & Conrad, 2010). These best practices describe the necessity of building community in an online course as well as the need for a variety of assignments. The discussion board assignments, final project, and chat room office hours described in this chapter were specifically designed in order to reflect GAISE guidelines as well as the best practices in the online environment.

27.4 Technology

Technology plays a vital role in an online course; however, the technology should be there to assist the course not be a hindrance. The subject matter of the course should be of primary importance for the students, not the technology used to deliver the course.

27.4.1 Hardware

For this course, the student needed a copy of the textbook and a copy of the course notes. The shell of the notes, a workbook, is a 120-page document with the examples, exercises, terms, and important concepts for the course; however, the examples and exercises are not completed. The students complete these examples and exercises with the instructor as they watch the tutorials. This allows for the students to concentrate on statistical understanding, rather than copying.

In order to understand the course design, it is necessary to understand the capabilities of Microsoft OneNote for the tablet PC. The program allows a user to include handwritten notes in a file by using a stylus. So, instead of working out a problem on a chalkboard, the instructor could work out the problems on the screen in OneNote. In OneNote, the color and width of the pen could be easily changed during the lecture. The shell of the notes that the students had was imported into OneNote allowing the instructor to write directly on the notes and then the instructor recorded video tutorials of the discussion. Microsoft OneNote was chosen rather than Microsoft Powerpoint because it allowed for a less restricted working space. For course creation the instructor used several pieces of hardware, including a microphone, webcam, and a Fujitsu T5010 tablet PC with a stylus pen and 4GB ram.

As for sound in the video tutorials, four sets of microphones were tested: the microphone built into the tablet PC laptop, the Azden WLX-PRO VHF wireless microphone, the Samson SWAM2SES N6 Airline Micro wireless ear set, and an H530 Logitech headset microphone. Two things should be considered when evaluating a microphone for online course use: the sound quality of the recording and the ease of use. Except for the Samson microphone, the other microphones had poor sound quality in the instructor's opinion and the breathing of the instructor was picked up on the recording. The Samson microphone did not pick up the breathing of the instructor and filtered out surrounding area noise making it the best option.

27.4.1.1 Content-Building Software

The content for the course was delivered in multiple ways using instructor built lessons and publisher supplied materials. For this course, SoftChalk was used to create a course website which covered 24 detailed lessons spanning 143 web pages and included complete topic explanations, 254 quiz questions, and 128 short instructor videos. The lessons also contained 22 activities including flash-based dynamic study tools written in SoftChalk, online applets, and exercise problems for the students to solve using StatCrunch. SoftChalk made it very easy to insert graphics, videos, sound clips, or web pages. It also made it possible to test the students on what they had learned and prepare flashcards or games to emphasize important topics. Grading these activities ensured that the student actually completed them and hopefully reached a higher level of engagement.

The lessons start with a list of about three to eight objectives that the student should master for that day. The lesson then steps the student through learning each of those objectives by first giving an explanation of why the objective is important and how it relates to the other material in the course. After the lesson objectives, there was a video (or videos) that explained the main concepts of those objectives. Following the video(s) there was usually either a quiz exercise, study tool such as flashcards, or a statistical applet for them to see the concept in action to help the student test their knowledge about what they had just learned.

For example, one of the first lessons of the course discusses “Measures of Center, Spread and Position.” On the first page of the lesson, a baseball data example is presented along with a brief discussion about why it would be important to quantify the measures of center, spread and position for this data set. The objectives of the day are then presented.

On the next page of the lesson, the first objective is discussed with a short tutorial which explains the definitions of the mean, median, and mode as well as finding the mean and median of two data sets. The students are then asked to complete a quiz where they have to find the mean and median of a data set in addition to matching the terms mean, median, and mode with their definitions. For the second objective, the students are asked to explore the effect of an outlier on the mean and median by playing with an applet designed by the publisher of the textbook and they are then quizzed on their findings. For the third objective, the students watch a short video that shows them how to compute the median from a stem and leaf plot from Minitab, a statistical software package. For the fourth objective, the measures of variance, standard deviation, and range are discussed in a short video and then the students are asked to compute these values for a data set as well as to predict the effect of an outlier on these measures. For the fifth objective, the empirical rule is explained in video and the students are asked to answer a question about the rule. The sixth objective includes videos that show how to compute the quartiles and how to use the quartiles to make, interpret, and read boxplots. The objective finishes with the students answering questions by comparing side-by-side boxplots. The last objective discusses output from Minitab and working with StatCrunch. Students watched a video by Webster West, the creator of StatCrunch and were asked to use StatCrunch to analyze a data set. The last page of the lesson reviews the important concepts that they have learned.

In addition to the SoftChalk lessons, the students were required to use MyStatLab for homework problems provided by the publisher for each lesson assignment. The homework problems accompanied the textbook, *Statistics: The Art and Science of Learning from Data* by Agresti and Franklin (2009). From an instructor’s point of view, the assignments were easy to select and assign and provided instant feedback to the students. But what was the experience from the student’s point of view? In the postsurvey, the students were asked how much time they spent working on the course (including everything related to the course: activities, quizzes, watching lectures, doing homework, studying) per day? The average amount of time spent on the course per day was 3.05 h. The standard deviation was 2.76 h. The data did have one outlier where the student said that they spent 15 h per day on the course, which seems

doubtful. If this point is removed, the average is 2.47 h with a standard deviation of 0.75 h. The students were also asked how many hours they spent on the course per week. The average time spent on the course was 11.64 h and the standard deviation was 5.36 h. The minimum number of hours per week was 3 and the maximum number of hours per week was 28.

Additionally, the students were asked in the presurvey how much of the homework they planned to complete. All but one of the responses said 100 % on the presurvey (one response said 80 %). On the postsurvey, the average response for the percent of homework completed was 97.5 % and the standard deviation was 3.628 %. The minimum was 90 % and the maximum was 100 %.

27.4.2 Computer/Video Screen Capture for Content Creation

Two forms of lecture capture software programs were used during the semester, Camtasia Relay and Jing!. Camtasia Relay was chosen because it was supported by the university and video storage was free but unlike the full version of Camtasia, it did not allow editing of the video beyond setting start and end times. Additionally, for Mac users an extra program called Flip4Mac had to be downloaded so they could watch the videos. Otherwise, Camtasia Relay was very easy to use. At the end of the semester, the students were asked what percentage of the videos they watched. Based on the students self reported usage, the average percentage of videos watched by the students was 91.73 %.

The other video capture program used was Jing! and it was selected for the student's projects because it was free and easily accessible online for students to use. The main limitation of free Jing! is that it only allows for 5 min recordings. A few students initially had issues understanding how the program worked and were resistant to learning a new software program. However, after they were pointed to the help tutorials on Jing!'s website, they were quickly able to make the software work. Afterward several students noted how easy it was to use and how they planned on using it in the future.

27.4.3 Interactive Communication Programs

During the semester, interaction was also encouraged between the students and between the instructor and the students. Several formats of interactive computer programs were used: email, the chat program in Sakai, the discussion board in Sakai, and the Elluminate software package. The students were sent list serv emails almost every day of the course reminding them of upcoming deadlines or giving additional instructions. The students were also encouraged to email questions about grades directly to the instructor and to post all questions about the content of the course and the administration of the course on the Q/A board.

The instructor initially thought not having a whiteboard in the Sakai chat function to answer questions would be limiting but it wasn't, instead the instructor made Camtasia Relay videos and posted it for the students. The Elluminate software allows instructors to conduct online office hours using video chat, text chat, or an interactive whiteboard. Additionally, it was possible for students to be polled about a concept or simply asked to raise their hand. Although the program had many capabilities, this also made it difficult to operate the program and teach at the same time. The initial plan for the online office hours for the course was to use the Sakai Chat room and then to transition to Elluminate. The transition, however, was not made due to complexity of the software. It felt that Elluminate had become the primary focus rather than learning the course material.

The discussion board in Sakai was used for two reasons: for a question and answer board and for a small group discussion board. Students were encouraged to post questions about the content of the course and general administration issues on the question and answer discussion board. The small group discussion board was used for discussion between randomly selected groups of about eight students. The students were asked to complete five activities during the 6-week semester. The first activity was for the students to introduce themselves to the group and then to reply to at least three other students' introductions. The second activity was for the group to select three articles from the internet that contained information about an experiment and/or survey. The students were then asked to identify various aspects of the study such as the explanatory and response variable and to discuss what aspects of the experiment/survey were good and what could be improved. The group then ranked each of the three surveys/experiments in terms of quality and adherence to the good survey/experimental protocol that they established. The third activity was for the students to conduct a lesson style called a Four Corner Debate that has the students debate a particular concept. The idea for a Four Corner Debate came from the talk by Michelle Everson and Jackie Miller at USCOTS 2011 (for more information on a Four Corner Debate visit this website http://www.educationworld.com/a_lesson/03/lp304-04.shtml). The concept for the debate was for students to consider issues about privacy and ethics as it relates to data collection and statistical analysis. Sometimes it is helpful for students to see other sides of an issue by not getting to pick the point of the view that they are arguing. So each student was told that in a few days a statement was going to be posted to the discussion board which they would need to debate. However, they had to pick their point of view before the statement was posted. The students had to pick if they "strongly agreed," "somewhat agreed," "somewhat disagreed," or "strongly disagreed" with the statement. Several days later the statement that "Data can only do good things in today's world," was posted. The students then had to support their point of view in respect to this statement. The fourth activity asked the students to complete a collaborative quiz on four questions with multiple parts about the sampling distribution of the sample proportion and the sample mean. The students were first asked to complete the assignment on their own and post their answers and then to work together as a group to complete a response from the whole group. This idea of the use of a collaborative online quiz came from a talk by Audbjorg Bjornsdottir and Ellen Gundlach who also

presented at USCOTS 2011. Only the final quiz responses from the entire group were graded and participation in building the team’s response to the assignment was a part of the grade. The last assignment was for the students to critique other students’ semester final project.

The instructor found grading the discussion board very time consuming for 67 students, particularly the second assignment. The students also resisted the group activities because they were uncomfortable coordinating with other students not in the same town or who didn’t respond in a timely manner. Additionally, for the fourth assignment there was very little discussion over the quiz answers since they didn’t want to point out that another student was wrong. In the future, the instructor plans to have students submit a group contract laying out each student’s responsibilities to help students feel more comfortable with the assignment.

The communication software and email were all used to improve interaction in the online course and to help build a sense of community. The students were asked in the pre- and post survey how important these technologies were to them and a few of the results are below (Table 27.1).

The students were also asked how important was interaction and how frequently they visited the Small Group Discussion Board (SGDB) and the class Q&A Discussion Board (QADB) (Tables 27.2 and 27.3).

Although this survey does not represent a random sample of students, it is interesting that the students preferred form of communication was still email.

Table 27.1 How important was the following with the instructor to you?

	Very (%)	Somewhat (%)	Minimally (%)	Not at all (%)
Interaction	55	35	10	0
Email	88	8	8	0
Discussion board	40	28	20	16
Online office hours	32	12	36	20

Table 27.2 How important was interaction with other classmates on the SGDB and QADB?

	Very (%)	Somewhat (%)	Minimally (%)	Not at all (%)
SGDB	8	24	40	28
QADB	12	12	44	32

Table 27.3 How frequently did you visit the SGDB and QADB?

	Every day (%)	Two or three times a week (%)	Several times a semester (%)	Only once per semester (%)	Did not participate (%)
SGDB	4	52	44	4	0
QADB	0	16	12	32	40

27.4.4 Grading

Determining how to setup high stakes testing in an online environment can be very difficult. The instructor needs to think about what type of assessments work the best at determining how well the students have learned the material, what type of mechanisms need to be in place to ensure that the students are who they say they are, and that the security and integrity of the exam itself remains protected.

For this course, the instructor determined that the best way to conduct high stakes testing was with an online proctored multiple choice test. All students had to begin their exam within 3 h of the first exam being started. The ordering of the questions and the answers was randomized for each student. The students were also directly proctored during the exam by an online test proctoring company called ProctorU. Before the exam, the students were encouraged to perform a system check of their computer to make sure that it would fully function with ProctorU's monitoring software. On the night of the exam, the students logged in to the ProctorU software and were greeted by a proctor in a video chat using a webcam. The students would then allow the proctor to see their computer screen so that whatever is on the computer screen is viewed by both the student and proctor as well. The proctor then asked to see the student's id and asked a few questions to ensure their identity. The company also took a picture of the student that could be used for later reference if needed.

For an instructor, setting up an exam time with ProctorU required completing a short Excel spreadsheet that included start and stop times, exam length, the date of the exam, the password of the exam, and if any special accommodations were needed. The instructor then setup the exam within the course management system and set a password for the test. The students would only find out the password after communicating with the proctor at ProctorU.

27.5 Conclusions

For conclusions, the specific course assessment will be given as well as a set of recommendations for teachers teaching the course for the first time.

27.5.1 Specific Course Evaluation

The course will be evaluated in two ways, the overall instructor evaluation and the overall course grades given. The overall instructor rating from student course evaluations was 4.42 out of 5. Table 27.4 shows the final grade distribution for the course.

Grades are a fairly limited source of assessment because they can be arbitrarily determined by the instructor. However, it does show that most students were successful in the course. The drop rate was 7.5 % and although this value is similar to other non-online courses taught by the instructor, it would be nice if the drop rate was smaller.

Table 27.4 Grade distribution for the course

Grade	A	A-	B+	B	B-	C+	C	D	E (failure)	Dropped
Number	19	5	9	10	4	4	6	2	3	5

27.5.2 Overall Assessment of Technology and Recommendations

SoftChalk was easy to learn for someone without a strong website HTML programming background. Additionally, its ability to add quizzes and other activities helped improve the learning experience. The tablet PC was a good tool allowing for quick graphics to be drawn for illustrating statistical concepts. Camtasia Relay recordings were easy to make and allowed for quick explanations of material to be presented to students. StatCrunch allowed for the students to collect and investigate their own data as encouraged by the GAISE guidelines. Finally, the SoftChalk lessons and MyStatLab allowed for immediate feedback while students were practicing working with statistical concepts. As for recommendations, don't assume that students will quickly pick up different software. Introductions to all forms of software used in the course should be provided to make the students more comfortable with the environment. Encourage communication through chat rooms and especially email since this is still their most comfortable form of communication. Software should be chosen to enhance a course and should take the back stage to the course material. Elluminate's software was cumbersome to use, whereas the other programs worked seamlessly in the background and aided in the learning of the material. Finally, more interference from the instructor to stimulate discussion and team work should be made. The discussion board assignments did not generate the sense of community that was their primary goal so students didn't work together well to make sure the work submitted was correct.

References

- Agresti, A., & Franklin, C. (2009). *Statistics: The art and science of learning from data* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., et al. (2005). GAISE College Report. Retrieved December 30, 2009, from <http://www.amstat.org/education/gaise/GAISECollege.htm>
- Bjornsdottir, A. & Gundlach, E. (2011, May 21). Rethinking assessment in the digital age. *United States Conference on Teaching Statistics*. Raleigh, NC.
- Boettcher, J., & Conrad, R.-M. (2010). *The online teaching survival guide: Simple and practical pedagogical tips*. San Francisco: Jossey-Bass.
- Everson, M. G., & Garfield, J. (2008). An innovative approach to teaching online statistics courses. *Technology Innovations in Statistics Education*, 2(1). Retrieved from <http://escholarship.org/uc/item/2v6124xr>

- Everson, M., & Miller, J. (2011, May 21). Using social media and technology to engage the next generation of statistics students. *United States Conference on Teaching Statistics 2011*. Raleigh, NC.
- Mills, J., & Raju, D. (2011). Teaching statistics online: A decade's review of the literature about what works. *Journal of Statistics Education*, 19(2). Retrieved from <http://www.amstat.org/publications/jse/v19n2/mills.html>
- Tudor, G. E. (2006). Teaching introductory statistics online—Satisfying the students. *Journal of Statistical Education*, 14(3). Retrieved from www.amstat.org/publications/jse/v14n3/tudor.html
- Utts, J., Sommer, B., Acredolo, C., Maher, M., & Matthews, H. (2003) A study comparing traditional and hybrid internet-based instruction in introductory statistics classes. *Journal of Statistical Education*, 11(3). Retrieved from <http://www.amstat.org/publications/jse/v11n3/utts.html>

Chapter 28

The Role of Technology in Indian Statistics Education: A Review

D.S. Hooda and B.K. Hooda

28.1 Introduction

Information technology has revolutionized the research and teaching methodology in last few decades. Its educational effectiveness is well established in many areas of research and teaching in statistics education. Statistics as a subject serves the dual purpose of efficiently collecting, recording, processing of observational data and making valid interpretations. It is a broad scientific discipline with theory and methods developed through use of mathematical tools and probability theory for making optimal decision under uncertainty. Originated to solve certain types of problems, the initial statistical methodology was based on common sense and convenience.

On the advancement in information technology, calculators and computers proved to be indispensable tools for analysis and interpretation of large data sets. Statistics education and practice these days is also supported by ever updating software packages. Initially known as a branch of applied mathematics, statistics has now emerged as mature discipline with its own philosophy and techniques. The role of technology in statistics education has been discussed by some statisticians, namely, Kish (1978), Prajneshu and Srivastava (1998), Ganchi (1999), and Sisodia (2007).

D.S. Hooda (✉)

Jaypee Institute of Engineering and Technology, A.B. Road, Raghogarh, Guna 473226, India
e-mail: ds_hooda@rediffmail.com

B.K. Hooda

Department of Mathematics and Statistics CCS, Haryana Agricultural University,
Hisar 125004, India
e-mail: bkhoda@gmail.com

Before the advent of high-speed computers, the subject of statistics, by and large, was considered as branch of mathematics applied to observational data. The success of classical statistics was because of a mathematically sound theory which was sufficiently realistic to handle a wide range of applications. But in the present scenario, the complicated mathematical arguments have been augmented with computing in the form of well-understood computer algorithms. Computer-intensive statistical methods have been developed in response to the challenges of making effective use of modern computational facilities.

Ironically, many statisticians in India and other developing countries are not well versed with their use and utility. This is mainly because of the flaws in statistics education and practice of statistics in these countries. The present study, therefore, is designed to give a brief account of Indian Statistical System. The role technology for effective teaching and learning statistics at different levels is discussed. Some suggestions and future strategies for improvement in statistics education in developing countries particularly in India are also discussed.

28.2 Backgrounds of Indian Statistical System Section

This collection and use of statistics for administrative purposes in India has a long history spread over many centuries. The *Arthashastra* and the *Ain-i-Akbari* mention the practice of numerical data collection for purposes of statecraft in ancient and medieval India. The Mughals had a system of collection and compilation of crop statistics to help them in land revenue collection. During the British period, consolidation efforts were made for the collection of socioeconomic data. But their system was restricted to a few specific fields like trade and commerce, selected industrial products, some basic crop statistics, and livestock. Just after independence in 1947, the system of data collection followed by the Britishers was found inadequate to meet the necessity of a strong database covering a variety of social and economical aspects. A very important step in this direction was the creation of the Directorate of National Sample Survey in 1950. Its aim was to collect essential statistics related to the socioeconomic conditions and agricultural production in India.

The Indian statistical system pertains to the collection, compilation, and dissemination of data relating to socioeconomic, agricultural, and industrial statistics in India. Under the Ministry of Agriculture (MOA), the Central Statistical Organization (CSO) and the National Sample Survey organization (NSSO) are some of the important agencies at the national level involved in collection, compilation, and dissemination of data. The CSO is mainly responsible for coordination of statistical activities as well as evolving and maintaining statistical standards. The NSSO has been a leading sample survey organization since its establishment and continues to conduct major multi-subject surveys that provide valuable data required by the policy makers.

The NSSO also conducts large-scale surveys at the national level and collects and disseminates information on different areas. The NSSO, under the scheme of

improvement of crop statistics, also provides technical guidance to the states in respect of the crop estimation surveys and performs sample checks to assess the quality of primary work done by the state agencies in area and crops estimation surveys. In addition to DOS and MOA, some other agencies like the state government departments, institutions and autonomous bodies and nongovernment organizations are also actively involved in the creation of large database.

28.3 Present Status of Statistics Education in India

Realizing the importance of statistics, some elementary topics are taught at 10+1 and 10+2 levels in school. At undergraduate level, statistics is taught as one paper in mathematics subject in most of the Indian universities; however, some other universities have introduced statistics as a separate subject. In general, statistics is taught in most of the Indian universities at the postgraduate level. Universities in some states offer UG as well as PG degrees in statistics. In order to impart training at postgraduate level, leading to M.Sc. and Ph.D. degrees, Calcutta University was the first to establish the departments of statistics in 1943. The several other universities also established separate departments of statistics and started postgraduate program in statistics.

During last few decades, statistics has penetrated into almost all sciences like biology, business, social, engineering, and medical including agriculture. Its wide and varied applications have led to the growth of many new branches, such as Industrial Statistics, Business Statistics, Biostatistics, and Agricultural Statistics. These branches have emerged as distinct entities with a bulk of statistical techniques specific to their application areas. It comprises the area of statistical science that deals directly with the problems of field experimentation and interpretation of results in agricultural sciences. Agricultural Statistics is the most important discipline regarding the training of the field research scientists, to help them in planning of their experiments and in the analysis of data and drawing inferences thereof. Statisticians working in Agricultural Universities/Institutions not only provide a strong technical support to other departments and disciplines for agriculture-related programs but also conduct their own research in Statistics.

In all Indian Agricultural Universities, Statistics is compulsory supporting subject in UG and PG programs of their disciplines. It is also taught as an independent subject leading to the formal degrees of M.Sc. and Ph.D. in Statistics. Generally, the teaching of Statistics particularly to agricultural background students is considered a tough job. This is mainly because of the fact that the agricultural students are not well trained in mathematical concepts and face difficulty in understanding the subject matter itself. Also, the teaching methodology in majority of these universities is traditional one of using chalk and blackboard. Further, most of the faculty members are on the verge of retirement and they are not well versed with the development in information technology software packages.

28.4 Technology in Statistical Education

Use of technology in the field of statistics education is commonly thought of in terms of computers and the associated statistical computing packages, which these days have become every statistician's indispensable working tool. Technology, especially in developing countries ranges from frequently used and low cost materials like chalk to the expensive and sophisticated computers and LCDs which are less used. Although there is no doubt that the impact of such new technology has caused changes in the nature of both statistical research and statistical practice, however, its effect on the teaching and learning of statistics varies considerably. Consequently, there is no uniformity in statistics education in Indian universities. Most of the universities start statistics at post graduate level and make use of the current statistical methodology based on probabilistic models developed for the analysis of small data sets, which is inadequate.

In order to make the statistical education system competitive, vibrant and useful, new upcoming technologies should be integrated with it and the statistics education curricula be made coherent. Recently, Indian Council of Agricultural Research (ICAR) constituted a Subject Matter Committee on the Broad Subject Matter Area (BSMA) of statistical science consisting of Agricultural Statistics, Biostatistics, and Computer Applications to update the curricula to meet the challenges of the millennium. This Committee prepared the new syllabi keeping in view the recommendations of the Third Deans' Committee Anonymous (1995) on Agricultural Education in India, syllabi of Agricultural Universities, and suggestions received from teachers of various State Agricultural Universities and Deemed Universities having a PG program in Statistical Science was in vogue.

The new syllabus prepared by this committee proved useful and helpful in unifying the teaching at the PG level to a great extent. The fast growing information technology needs to be integrated with the classroom teaching to provide adequate support to the teaching of the subject. The theory needs to be supported and explained by live practical examples through computers. So students should be trained in the use of computers and application of various standard statistical software packages for analysis of data using different statistical techniques. The network technology could be advantageously exploited to update the knowledge of the statisticians working in remote regions of the country through distance education program. Such program may consist of supplying lecture notes and other reading material through Internet.

Application of computer-based graphics and methods of data processing has opened ways for the development of new statistical methods. Development of the widely used statistical methods like generalized linear models, nonlinear regression, Neural Networks, Cross Validation, Monte Carlo simulations, Jackknife and Bootstrapping, and Image Analysis would have been unthinkable without the availability of high-speed computers.

From an instructional perspective, especially while teaching computer-intensive methods, teachers sometimes struggle to find ways to demonstrate to the expectations

of students. It is often difficult to demonstrate abstract concepts and consequently to tell students how they would be assessed. Teachers' expectations can be made clear when technology is used to post concrete examples. When such examples are made available to students, they can better perform to meet the teachers' expectations regarding instructional and assessment goals. Just as students need time to adapt to new technologies in the context of instruction, in the same way teachers will need assistance in learning how such technologies can be used to facilitate instruction and assessment.

The use of graph theory in describing the nature of data, choosing appropriate model, and reporting and interpreting the results of statistical analysis is well known in statistical literature. After development of high-speed computers and possibilities of viewing high dimensional data through parallel coordinates and data reduction by canonical coordinates and principal components, graphical analysis is becoming a valuable tool in discovering patterns in data. Automatic recognition of various aspects of human face can now be accomplished with computer and image processing technology.

On the other hand, wide applicability of computer-intensive methods in statistical education and practice has made significant progress in statistical education in developed and advanced countries. For instance, many problems in statistical inference can be solved in seconds, which otherwise supposed to take days or months to solve manually or using simple calculators. Such problems may require evaluations of integrals and solution of equations having no explicit solutions but can be solved by numerical evaluation. In particular, the Jackknife Method helps in reducing bias where unbiased estimates are not available.

It consists of leaving, say, one observation at a time and calculating the estimate from the remaining values. Averaging then all the estimates gives an estimate that has made smaller bias.

The bootstrap is another computer-intensive method known as re-sampling method which provides estimates of standard errors especially when they are not easily obtainable. It takes repeated samples from the empirical distribution function of the obtained sample. The statistics of interest is then evaluated from each sample. Thousands of such samples are taken, the statistic is computed, and then standard errors and confidence intervals can be obtained using the empirical distributions. The novel techniques of data mining, Jackknifing, bootstrap, pattern recognition, and image analysis are essentially computer intensive and frequently used techniques in advanced countries, and statisticians in developing countries are still facing difficulty in using such techniques.

28.5 Strategies for Improvements in Statistics Education

New technologies like e-commerce, e-business, and e-learning are emerging with the advent of multimedia and Internet technologies. There is a need to include elements of these topics suitably in the curriculum to equip the students with all

modern computing tools for effective and hi-tech decision making during their professional career in future. These technologies provide an advanced area of educational technology that can be applied to statistical education too. Tutorial software, including 3D graphics and animation effects, may be developed by the teachers for self-guided learning by students and to understand the otherwise complicated concepts and methodology of statistics. The students may complete such sessions at their own place by sitting on their personal computers at their homes, hostels, etc. For making this an effective technique, students should be encouraged to keep their own Laptops or PCs.

Further, the teachers and students can exploit the intranet facilities existing on the campus for effective interaction. Teachers can provide handouts, assignments and make various announcements for students through the local area network system of the institute. They can also make available the solutions to the question papers and communicate the comments to the students' assignments through the LAN. This is supposed to help the students to judge their performance themselves. Similarly, students can make queries to the teachers in case of any doubt and submit their assignments using the LAN facilities. Also, the new state-of-the-art technology should be adopted immediately by the universities/institutions so as to impart training to students as well as the faculty members on the latest computing equipment for improvement in the statistics education.

28.6 Conclusions

The new challenges in statistics education are never trivial but offer opportunities inherent to the richness of statistics as a discipline and a servant. The external challenges current in India are many including:

- University structures that tend to encourage competition rather cooperation between faculties and thus tend not to reward good service teaching.
- The Juxtaposition of a general decline to reward good service in teaching and in quantitative preparation versus increased quantitative needs of business and industry.

The awareness in statistical profession of the importance of statistics education and commitment of teaching community to their students and good applications of technologies in teaching of statistics have laid excellent foundations for the future.

References

- Anonymous. (1995). *Report of third deans' committee on agricultural education in India*. New Delhi: An ICAR.
- Ganchi, D. A. (1999). The cyber-age, education and the teaching program. *University News*, 25, 3-6.

- Kish, L. (1978). Chance, statistics, and statisticians. *Journal of the American Statistical Association*, 73(361), 1–6.
- Prajneshu, & Srivastava, R. (1998). Standardization and monitoring of teaching in agricultural statistics at M.Sc. and Ph.D. levels in Agricultural Universities. In *Proceedings of the XII-National Conference of Agricultural Research Statisticians* (pp. 187–193).
- Sisodia, B. V. S. (2007). Teaching of statistics to post graduate students of agriculture and allied sciences. In *Discussion Papers and Proceedings of Fifth National Conference of Agricultural Research Statisticians* (pp. 124–125).

Chapter 29

Building Up the Box Plot as a Tool for Representing and Structuring Data Distributions: An Instructional Effort Using *Tinkerplots* and Evidence of Students' Reasoning

Luis Saldanha and Michael McAllister

29.1 Introduction and Background

The box plot is a powerful reductive graphical inscription for representing and comparing data distributions, allowing the user to attend to key features visually, such as summary of center and spread of the data, while obscuring individual values. In the USA, box plots have made their way into mathematics curricula in Grades 6–8. In recent years, however, several researchers have investigated and documented students' challenges in understanding and using box plots as a tool for representing and comparing data distributions. According to Bakker, Biehler, and Konold (2004), young students' sources of difficulty include the fact that box plots mask individual data values and convey only aggregate features of a data set. More specifically, box plots display densities rather than the more intuitive or familiar frequencies. In addition, partitioning data sets into quartiles and thinking of the median as a measure of center appear to be far less intuitive or sensible for students than we might assume. One of Bakker et al.'s salient instructional recommendations is that box plots be introduced in combination with dots plots of data sets in appropriate ways, such as overlaying them on the latter before obscuring the individual cases. Similarly, in a study of Grade 11 students' ability to use box plots to make group comparisons and inferences, Pfannkuch (2007) suggested that “keeping data under the box plot for as long as possible” (p. 164) may be a factor in improving students' inferential reasoning. Garfield and Ben-Zvi (2010) have integrated these suggestions into their development of lesson sequences for reasoning with box plots and using them as a tool for comparing groups. Konold (2007) seems to have taken these suggestions into

L. Saldanha (✉)

Université du Québec à Montréal, Montréal, Québec, Canada

e-mail: saldanha.luis@uqam.ca

M. McAllister

Arizona State University, Tempe, AZ, USA

e-mail: memcalli@asu.edu

consideration in his design of the *TinkerPlots* software (Konold & Miller, 2005). *Tinkerplots* is a constructive and dynamically linked data analysis environment designed for students in Grades 6–9. The software was borne out of a research and development effort that drew on research findings summarized by Bakker et al. (2004). *TinkerPlots* provides students with a number of intuitive precursors of the box plot, and it allows great flexibility in building up and overlaying box plots on graphs that display individual cases, and in obscuring both types of inscriptions as needed.

A salient instructional implication that we draw from this literature and product developments is that it may be productive to “build up” students’ understanding of box plots by engaging them in the activity of organizing and representing data sets in ways they might see as natural precursors of box plots. Our chapter describes an instructional effort in that direction and presents evidence of students’ thinking that emerged from their engagement with this instruction.

29.2 Participants and Their Prior Knowledge

Six students chosen from a seventh-grade class in a suburban middle school in the southwest United States participated in a group teaching experiment involving a sequence of 8 after-school lessons that engaged them in analysis of univariate data sets within the *TinkerPlots* software environment (Konold & Miller, 2005). Students had been exposed to some basic statistical ideas and graphs within curricular units of their previous coursework. Their responses to pretest questions revealed some of their prior statistical skills and knowledge: students were able to construct and read frequency histograms and dot plots; they were familiar with terms like “mean” and “median,” but had limited understandings of the meaning and significance of the former.

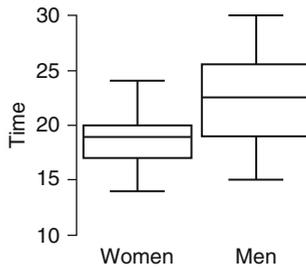
With specific regard to box plots, Fig. 29.1 displays a pretest scenario¹ and three accompanying clusters of questions used to query students’ understanding of box plots. Table 29.1 shows the features identified in students’ responses to Question (i) of this scenario and the number of students who identified each of them.² The table indicates that representation of extreme values, median, and range were salient aspects of box plots for students. A striking result is that only one student identified quartiles.³ Although not indicated in the table, two students did implicitly allude to the second and third quartiles as a unified region representing the “majority” or “bulk” of data values. One student remarked that the box plot does not indicate the number of data values being represented.

¹ Scenario and box plots were retrieved from [http://onlinestatbook.com/chapter2/Box plots.html](http://onlinestatbook.com/chapter2/Box%20plots.html) on April 28, 2011.

² The results for Question (ii) were nearly identical to those for Question (i) and are therefore not displayed here.

³ We follow Tukey’s (1977) usage and refer to “hinges” as the numerical values that delineate the boundaries of the boxes and whiskers, and to “quartiles” as the intervals constituted by those boxes and whiskers.

An experiment was conducted with a group of women and a group of men. The participants in each group were presented with a sheet containing 30 colored circles. Their task was to name the colors of all 30 circles as quickly as possible; their times (in seconds) were recorded. The box plots of the data collected for each gender are shown below.



Decode Boxplots

- i) Write down all the information that the **Women's** box plot tells us about the women's performance in the experiment.
- ii) Write down all the information that the **Men's** box plot tells us about the men's performance in the experiment.

Interpret Boxplots

- iii) Describe how the men and women compare overall in their speed of identifying colors.
- iv) How do the two groups compare with regard to the variability of their respective times?

Create Data Set

- v) Twelve (12) women participated in this experiment. Make up a set of data values for the women's times (in seconds) that could be represented by the women's box plot. Draw your values as points on the plot.

Fig. 29.1 Pretest items used to query students' prior knowledge of box plots

Table 29.1 Frequency of students' identification of box plot features in pretest Question (i)

Features identified					
	Max and min	Median	Quartiles	Interquartile range	Range
Frequency	6	6	1	2	4

All six students' responses to Question (iii) concluded that women performed faster than men in the experiment described in the problem scenario. Moreover, five students justified their conclusions by explicitly citing appropriate supporting features of the box plots. Some indicated that either or both the women's median and middle 50 % were below those of the men, respectively; one student also said that the women's range was smaller than the men's, while another said that the men's data were more spread out than the women's. Regarding Question (iv), only two

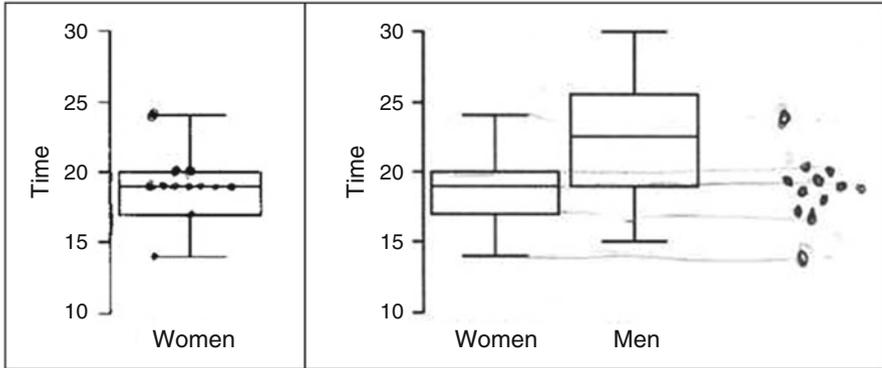


Fig. 29.2 Two types of student responses to pretest Question (v)

students understood the question. Both of those students associated variability with range: one said “the men have a larger range than women,” and the other said “men had a broader range so their times varied more.” Students’ responses to Question (v) are arguably most indicative of the limits of the imagery that box plots evoked for them. Three of the six students created a data set *all of whose values* coincided only with the box plot’s hinges, concentrated at the median and containing no data within the quartiles (Fig. 29.2, left panel), despite having mentioned quartiles or interquartile range in their responses to the preceding questions. The other three students created data sets whose values were concentrated within only the second and/or third quartiles (Fig. 29.2, right panel), instead of being dispersed throughout all four quartiles. This last is consistent with their thinking that only the boxes, and not a box plot’s whiskers, represent quartiles of a data set.

Overall, the results of these pretest questions indicated that students were able to use box plots globally as an inscription for making valid group comparisons. However, their focus on the salient global features indicated in Table 29.1—which was almost to the exclusion of attention to quartiles—together with their responses to Question (v) also suggested that their understanding of box plots was somewhat fragmented and incomplete. There was little evidence of an understanding that box plots partition a data set into ordered quarters. Additionally, students had largely impoverished images of how a data set represented by a box plot might be distributed.

29.3 Instructional Method and Sample Activities

In light of the prior research and the abovementioned indications about our students’ understandings of box plots, we designed a sequence of instructional activities to support their understanding the box plot as a tool for representing and measuring patterns of dispersion within univariate data sets.

As a preamble to the instructional sequence, students first spent two lessons exploring a data set consisting of battery life-span values (Bright et al., 2003). Students discussed aspects of the data that were salient to them and posed questions about the diversity of values in the set. They then displayed the data in a graph of their choosing in order to answer a question designed to orient their attention to the data set's dispersion and variability: *Are the life-span values in this data set, on the whole, very similar to each other or do they differ a lot?* This exploration concluded with a class discussion about students' graphs in relation to this question. In preparation for Lessons 3–5—those constituting the sequence on which we report here—the preamble activities culminated with a tutorial of the *TinkerPlots* software (Konold & Miller, 2005) in which students experimented with its basic features and learned to create histograms and dot plots of the battery life-spans' data set.

Each of Lessons 3–5 lasted approximately 65 min and was held on a separate day within a period of 1 week. The instructional sequence aimed to promote students' reasoning about patterns of dispersion within univariate data distributions by focusing their attention and reflection on two attributes: the *location* of portions of the data along a continuum of values and the *spread* of those portions. The activities moved to achieve this by having students first employ *TinkerPlots*' *divider* and *ruler* tools (in Lesson 3), and then the *percentile hat* tool (in Lessons 4–5) to organize ordered dot plots of the battery life-spans data set in ways intended to highlight the abovementioned attributes. Activities were presented in the form of structured worksheets that guided students through a sequence of actions to perform on the data set in *TinkerPlots*, followed by reflection questions about the results of those actions. Students worked the activities in pairs on a computer and responded individually to the reflection questions. This was typically followed by discussions of students' responses and their reasoning with regard to the questions.

Figure 29.3 shows a representative excerpt from the activity worksheets of Lesson 4 that engaged students in using the percentile hat and ruler tools. Figure 29.4 displays

Exploring the Spread Out-ness of Data Sets	
<p>In this activity we will use the <i>Percentile Hat</i> tool to organize data by highlighting the first, second, third, and fourth lowest 25% of the data values (splitting the data set into quarters).</p> <p>Our goal is to understand how a data set is distributed—where each of these portions of the data values is <i>located</i> and how <i>spread out</i> is each portion.</p> <p><i>Important:</i> Please answer each question (do your best) and then discuss with your partners.</p> <ul style="list-style-type: none"> • Open TinkerPlots file named <i>Batteries_AlwaysReady</i>. • Make a stacked-dot plot of the data set, showing the values fully separated. • Use the <i>Percentile Hat</i> tool to highlight the <u>lowest 25%</u> of the data values. <p><i>Question 1:</i> What does the percentile hat show about where the lowest 25% of the data values are located?</p> <p><i>Question 2:</i> How <i>spread out</i> is this portion of the data set? Use the <i>Ruler</i> tool to help answer this question.</p> <p>I know this because:</p>	<ul style="list-style-type: none"> • Make a copy of your plot by choosing <i>Duplicate Plot</i> from the <i>Object</i> menu. Drag this copy below your original plot. • Use the <i>Percentile Hat</i> tool to highlight the <u>second lowest 25%</u> of the data values. <p><i>Question 3:</i> What does the percentile hat show about where the second lowest 25% of the data values are located?</p> <p><i>Question 4:</i> How <i>spread out</i> is this portion of the data set? Use the <i>Ruler</i> tool to help answer this question.</p> <p><i>Question 5:</i> How do these two portions of the data (first and second lowest 25%) compare in their spread out-ness? Please explain how you know this.</p>

Fig. 29.3 Excerpts from the activity sequence of Lesson 4

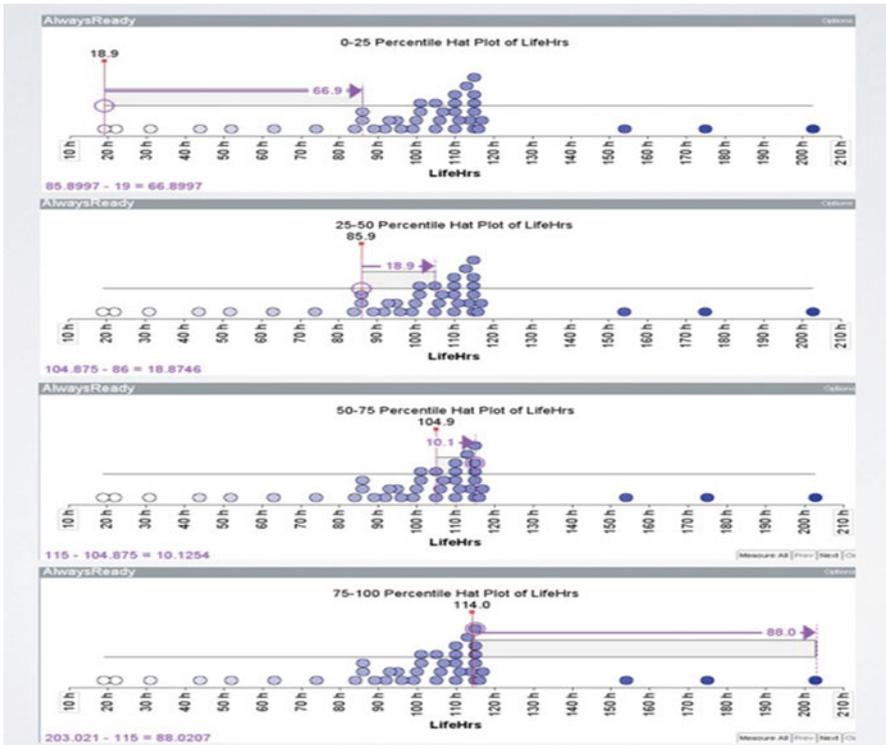
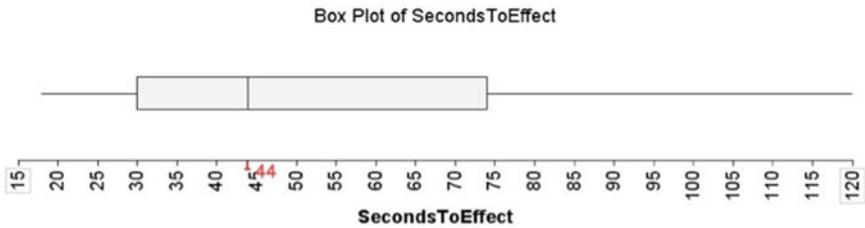


Fig. 29.4 A sequence of four dot plots of the battery life-spans data, each with one quartile highlighted by a *TinkerPlots* percentile hat and its spread indicated by the ruler tool

a screenshot of the associated product of students' work in *TinkerPlots*, which was to form the basis of their answers to the worksheet reflection questions. The percentile hat tool gives the user precise control in partitioning a data set by dragging the edges of a hat's "crown" to highlight any desired portion of the data displayed as a dot plot. The ruler tool allows the user to measure the length of any portion of a data set by dragging an arrow icon from any initial point to any end point along a horizontal or vertical direction and returning the length of the resulting segment. The worksheet (Fig. 29.3) guided students' activity of methodically using the percentile hat tool to partition the data set into quartiles, moving in order from the lowest to the highest, and measuring the length of each quartile with the ruler tool. For each quartile so highlighted (Fig. 29.4), students answered questions about the location and spread of that portion and about how adjacent quartiles compared in their spreads.

The systematic use of *TinkerPlots*' divider, percentile hat, and ruler tools to partition and highlight portions of a data set in the sequence of activities across Lessons 3 and 4 was intended to provide students with an imagistic basis for how they might organize and structure data sets when comparing them, and for orienting their attention to the location and the spread of portions of a data set when doing so. The particular sequence of actions and questions entailed in the worksheet of Lesson 4 (Fig. 29.3) was specifically intended to build toward the box plot as a conventional inscription that

a) Guess what the “hidden” data set might look like:



b) What does this box plot tell us about how the data values might be distributed?

Fig. 29.5 A culminating task of the instructional sequence. (a) Guess what the “hidden” data set might look like. (b) What does this box plot tell us about how the data values might be distributed?

highlights the partitioning of a data set into ordered quarters (Konold, 2007), so that students would see box plots as emerging naturally from their activity of having organized a data set in the way that they had. Indeed, this intention provided the entry point for Lesson 5, which focused on explicating and solidifying the connection between students’ prior activity and box plots. Here, the instructor first recapped what students had done and created in Lesson 4, and then introduced a box plot of the data set as the culmination of the sequence of percentile hat plots shown in Fig. 29.4. Discussions focused on explaining how each of the quartiles and their boundaries shown in their graphs (Fig. 29.3) corresponded to the box plot’s whiskers or boxes, and hinges, respectively. The aim was for students to see the box plot as a very natural, but small, extension of their ordered sequence of 25th-percentile hat plots.

Figure 29.5 displays a culminating task of the instructional sequence that we will discuss in this report. Students had to (a) create a hypothetical data set that could be represented by a given box plot, and (b) explain what the box plot indicates about how a represented data set might be distributed.⁴

29.4 Data Corpus and Analysis

The data corpus collected over the course of the entire teaching experiment includes video recordings of group discussions around the instructional task sequence and questions, students’ written responses on these and pre- and posttest questions, and video recordings of individual interviews conducted with students at the end of the experiment. However, in addition to the pretest responses already discussed in an earlier section, our report focuses only on students’ responses to the questions in the culminating task of Lesson 5 (see Fig. 29.5).

⁴The task was designed to mimic Question (v) of the pretest, so that we might assess students’ thinking at post-experiment.

Table 29.2 Salient features of students' data sets created in the culminating task, part (a)

Features of students' data sets						
Student	Size	Correct extreme values?	No. of values at each hinge	No. of values within each quartile	Dispersion within quartiles	Diverse distribution?
S1	26	Yes	1 or 2	6	Uniform	Yes
S2	28	No	1 or 2	7	Uniform	Yes
S3	20	Yes	1 or 2	5	Clustered near hinges	Yes
S4	32	Yes	1	7	Uniform	Yes
S5	9	Yes	1	0 or 2	Uniform	No
S6	9	Yes	1	1	Uniform	No

Table 29.3 Frequency of students' identification of features in the culminating task, part (b)

Features identified					
	Box plot divides data into quarters	Box plot indicates location of quartiles	Quartile length indicates spread	Range indicates spread	Median indicates center
Frequency	4	2	4	3	1

Each of the authors first independently examined students' responses to the culminating task for evidence of their plausible imagery and conceptions related to a data collection's patterns of dispersion. We then compared each of our documented evidences and created an initial set of codes that focused on capturing salient aspects of the students' created data sets (part (a)) and their responses to part (b) of the task. Through a process of successive applications of these initial codes to the students' work and responses, we refined the codes until they converged to the set of features described by the column headings in Tables 29.2 and 29.3. This convergence was driven partly by our interest in providing a snapshot assessment of the richness of students' imagery when creating a data distribution under the structuring constraints of a given box plot, and partly by consideration of the features identified in students' pretest responses (see Table 29.1 and Fig. 29.1) for the purpose of comparison.

29.5 Results and Discussion

29.5.1 *Culminating Task (a): Guess the Hidden Data Set*

Table 29.2 summarizes the results of our analysis of students' responses to part (a) of the culminating task (Fig. 29.5). Four of the six students created data sets containing between 20 and 32 values, while two had relatively small sets containing

only 9 values (Column 1). We will speak to the potential significance of this later in this section. All except one student chose appropriate minimum and maximum values (Column 2). The exception (S2) chose only a correct minimum value; this was possibly just a careless omission, given the lack of such error on pretest Question (v).

A salient feature of the students' data sets was the small number of values placed at the box plot's hinges (Column 3), particularly in the larger sets. Another salient feature is shown by the entries in Column 4 of the table: five of the students created a data set that contained values within all four quartiles, whereas the exception (S5) placed no values in the first and fourth quartiles. Figure 29.6 displays the work of one of these five students (S1) on this part of the culminating task. These last two features are in contrast with the results of pretest Question (v): as reported in an earlier section, half of the students created a data set of the type shown in the left panel of Fig. 29.2 (also made by S1), involving no values within the quartiles. The others created data sets whose values were concentrated largely within the second and third quartiles, as if whiskers did not represent quartiles for them.

As indicated by the entries in Column 5 of Table 29.2, all but one student's data sets had their values distributed fairly uniformly within each of the quartiles. Figure 29.6 also exemplifies this type of data set. The exception (S3) is shown in Fig. 29.7, wherein the values tend to be clustered near the box plot's hinges.

Each entry in the last column of Table 29.2 indicates our assessment of whether that student's data set was relatively diverse. We view this as a summary assessment of a data set's distributional richness, relative to the structure of the given box plot. Thus, in order for a data set to be assessed as diverse in the context of this task, it

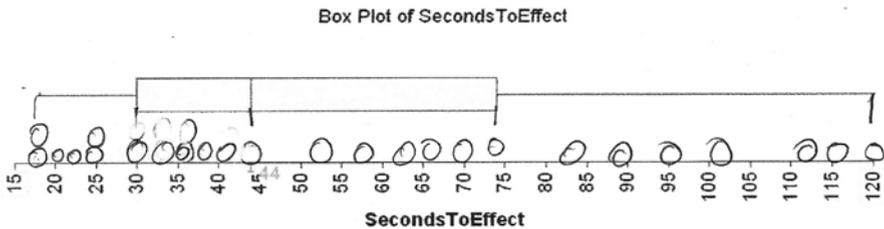


Fig. 29.6 S1's data set from the culminating task of Lesson 5

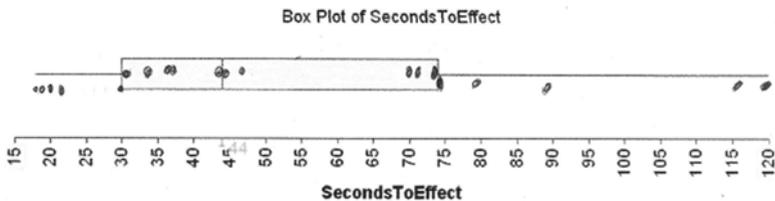


Fig. 29.7 S3's data set having a nonuniform distribution of values within quartiles

must be large enough to provide opportunity for creating a variety of data values, it should contain few or no repeated values, and it should have a distributional structure commensurate with the given box plot. We considered four students' data sets to be diverse according to this definition. The data sets produced by students S5 and S6 were both too small to exhibit a rich distributional structure, and in the case of S5 did not fully adhere to the box plot's structure. These diversity assessments also suggest that four of the students enriched their imagery of data distributions from pre- to post-instruction. This is likely a consequence of their engagement with our instructional sequence.⁵

29.5.2 *Culminating Task (b): What Does the Box Plot Indicate about How Data Is Distributed?*

Table 29.3 summarizes features identified in students' responses to part (b) of the culminating task (Fig. 29.5) and it shows the number of students who identified each of them.

Table 29.3 indicates that the most salient features, overall, for students were that the box plot partitions the data set into quarters and that a quartile's length indicates how spread out is that portion of the data. Regarding the former, three students specified that these were equal percentages independent of the number of values in each. Regarding the latter, all students who mentioned it also indicated that length and spread are in a direct relationship.⁶ The frequency of the two features in Columns 1 and 3 are in contrast with the results of the pretest, which indicated that quartiles were largely outside of students' attention. While the results in Columns 1 and 3 of Table 29.3 indicate increased attention to quartiles, particularly to their length as an indicator of the spread of data, the frequency count in Column 2 also suggests that the location of quartiles may not have become as salient to students as we intended. Regarding this last point, we have two conjectures. One conjecture is that our instructional sequence's strong focus on the relationship between quartile length and spread of data may have inadvertently overshadowed the idea of a quartile's location in students' minds. Another conjecture is based on evidence from the group discussions and individual interviews that we cannot present here due to space limitations: a number of students exhibited persistent difficulty in teasing apart a quartile's location from its length, as though the latter implied the former and was thus not viewed by students as a separate attribute. These conjectures and the relationship between them will be explored in a future report that draws on our extended data corpus and on a subsequent whole-class teaching experiment that we conducted with a different group of students.

⁵These students were not studying statistics in their regular class during the semester in which they participated in the teaching experiment. It is therefore implausible that this enrichment was due to their regular school instruction.

⁶This information is not indicated in the table.

29.6 Summary

We have reported on part of a teaching experiment that engaged six 7th-grade students with a sequence of activities in which they used *TinkerPlots* to organize a data set in ways intended to build toward the structure of box plots—by highlighting and comparing the location and spread of ordered quarters of the data set. Our comparison of students' responses on pretest and culminating task questions suggests that their imagery of data distributions and of box plots as representing quartiles and spread was consequently enriched, but that location of quartiles was not as salient for them as we had intended.

Acknowledgment This report is based upon work supported by the National Science Foundation under Grant No. 0953987. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about box plots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: international association for statistical education 2004 roundtable* (pp. 163–173). Voorburg, The Netherlands: International Statistical Institute.
- Bright, G. W., Brewer, W., McClain, K., & Mooney, E. S. (2003). *Navigating through data analysis in grades 6-8*. Reston, VA: National Council of Teachers of Mathematics.
- Garfield, J., & Ben-Zvi, D. (2010). *Developing students' statistical reasoning: Connecting research and teaching practice*. Dordrecht, The Netherlands: Springer.
- Konold, C. (2007). Designing a data analysis tool for learners. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 267–292). New York: Lawrence Erlbaum Associates.
- Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration*. Emeryville, CA: Key Curriculum Press.
- Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education*, 2(3), 149–167.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Part V
Statistics Teachers and Teaching

Chapter 30

Integrated Reasoning About Statistical Variation: Secondary Teachers' Development of Foundational Understandings

Susan A. Peters and Ksenia S. Kopeikin

30.1 Introduction and Background

The explosion of digital data in today's technologically driven world necessitates graduating statistically savvy students. Policy makers and educators have addressed this need by recommending the significant study of statistics for all students (e.g., NCTM, 2000). High-quality teachers are important for students to reach curricular goals. Research suggests that the largest school factor affecting student achievement is teacher quality (e.g., Hanushek, Kain, & Rivkin, 1998), and an important aspect of teacher quality is knowledge. Recent studies confirming the positive effects of teacher knowledge on student achievement further underscore the importance of teachers being knowledgeable about the content they teach (e.g., Hill, Rowan, & Ball, 2005).

Expository literature presents a prevalent view that many teachers have not had sufficient statistics learning experiences to develop requisite knowledge of statistics for teaching (e.g., Shaughnessy, 2007). These views are supported by research that reveals teachers' difficulties with understanding fundamental statistical concepts such as average or variation (Batanero, Burrill, & Reading, 2011). Teachers' difficulties with statistics content precipitated a joint ICMI/IASE study (Batanero et al., 2011). Much of the work associated with the study detailed the limited statistical

S.A. Peters (✉)
University of Louisville, Louisville, KY, USA
e-mail: s.peters@louisville.edu

K.S. Kopeikin
Suffolk University, Boston, MA, USA
e-mail: kkopeikin@suffolk.edu

knowledge that a majority of teachers expressed. Novel programs offering promise for developing teachers' knowledge were also examined.

Rather than focus on what teachers do not know or design a program that might be successful in developing teacher knowledge, the larger study from which this chapter emerges focuses on teachers with robust understandings of statistical content and explores the activities and experiences that led to development of those understandings. This chapter focuses on teachers' learning of the foundational concept of variation—learning that arguably parallels the general learning of statistics—to address the research question: for secondary statistics teachers who exhibit integrated reasoning about variation, what are the factors that contributed to their current understandings of variation as reflected in their perceptions and recollections?

30.2 Theoretical and Conceptual Frameworks

Design, data centric, and modeling perspectives form the basis of the perspectives framework employed by this study to consider the nature of teachers' reasoning about variation. Reasoning about variation from the design perspective includes identifying the nature of and potential sources of variation and considering strategies to control variation. Reasoning from the data-centric perspective entails measuring, describing, and representing variation while exploring distributions and relationships among data and variables. Reasoning about variation from the modeling perspective encompasses modeling data or data characteristics to infer relationships among data and variables. Indicators of reasoning from each perspective emerged from analysis of teachers' responses to variation-related tasks (Peters, 2009), research related to variation (e.g., delMas & Liu, 2005), and expositions on understanding variation (e.g., Garfield & Ben-Zvi, 2005). Analysis of teachers' reasoning from the larger study led to identifying four elements that transcend perspectives: variational disposition, variability in data for contextual variables, variability and relationships among data and variables, and effects of sample size on variability. *Integrated reasoning about variation* incorporates elements within one or more perspectives or multiple elements across perspectives. The complete list of indicators for each element from each perspective and description of the analysis that led to their identification appear in Peters (2009, 2011).

Transformative learning theory (Mezirow, 2000) serves as this study's theoretical frame. Grounded in constructivist assumptions, an overarching premise of the theory is that adult learning differs from student learning. Reflection is an important aspect of learning, but most students do not have the life experiences to critically evaluate the assumptions and beliefs upon which their knowledge is built. Students' reflections tend to focus on questions of what and how—the content and processes of educational studies that are important for constructing understanding if new ideas are encountered or to enhance current understanding. Many adults have the capacity to also reflect on the premises behind content and processes to question the importance, validity, or utility of knowledge. Reflection on premises—*critical*

reflection—is a key component of transformative learning. Critical reflection typically begins with events that *trigger dilemmas*—conflicts from internal or external stimuli that signal dissatisfaction with knowledge. An individual may become aware of previously implicit assumptions and assess those assumptions through critical reflection. Upon critical reflection, the individual may explore options for new roles and actions by engaging in rational discourse to develop and act on a plan to resolve the dilemma(s). In *rational discourse*, the individual engages in discourse with self or others to question assumptions, weigh evidence, and assess justifications to resolve dilemmas while searching for understanding. The main elements of critical reflection, rational discourse, and action are used to frame teachers' experiences.

30.3 Methods

This study uses phenomenological methods (Moustakas, 1994) to investigate the phenomenon of developing understandings of variation for integrated reasoning about variation. To ensure the highest probability of finding teachers who experienced the phenomenon, the lead author developed explicit criteria for teachers' inclusion in the study. Due to their wealth of background experiences, AP Statistics teacher-leaders were consulted because they were more likely to exhibit integrated reasoning about variation than a random sample of teachers. School students receive college credit for introductory statistics if they successfully complete an AP Statistics course and examination. To consider diverse experiences, teachers who were selected for participation: differed in the number of years they taught statistics, displayed a variety of educational backgrounds and course-related statistics experiences, encountered varied professional development learning and teaching experiences and informal experiences with statistics, exhibited leadership through a variety of venues, and differed according to other characteristics such as gender. The resulting purposeful sample consisted of 16 secondary statistics teacher-leaders from 14 states in the United States.

The larger study explored factors that contributed to teachers' development of robust understandings of variation for the five teachers who exhibited reasoning consistent with robust understanding (Peters, 2009). Processes used to identify these teachers appear in Peters (2009, 2011). The same processes led to identifying five more teachers who exhibited integrated reasoning about variation. This study considers the learning of these ten teachers.

30.3.1 Data and Instruments

Phenomenological studies often use semistructured interviews to elicit feelings about and experiences with the studied phenomenon. Participants retrospectively recall their experiences and feelings through self-report. Because of the reliance on memories and concerns related to the accuracy of those memories, this study

employed event history calendars (EHCs), critical incident reports, and two 90- to 120-min semistructured interviews to enhance teachers' abilities to recall and report learning events accurately.

EHCs enable individuals to reconstruct past events and experiences by using cues for significant past events (Martyn & Belli, 2002). The calendar is a table with timing cues for columns and significant events for rows. Teachers recorded personal information, including times when they were studying statistics and teaching mathematics or statistics, and events and experiences related to their statistics education, including events related to AP Statistics and to professional development. For each experience, teachers recorded details of the event and people, places, and feelings associated with the experience. See Peters (2009) for an example of an EHC and examples of other study instruments.

Critical incidents (CIs) are unique events that are significant to individuals. CIs "highlight particular, concrete, and contextually specific aspects of people's experiences" (Brookfield, 1990, p. 180) and can serve as a window for inferring people's assumptions and beliefs. Teachers recounted two CIs related to their formal or informal study of variation or statistics—one positive and one negative—to focus on salient experiences.

Interviews allow participants to reconstruct the finer details of experiences (Seidman, 2006). Two interviews reconstructed details of experiences listed in teachers' EHCs and CIs. Prior to conducting each teacher's first experience-related interview (Int I), the lead author used the teacher's EHC and CIs to determine the temporal positioning of educational experiences, become familiar with the teacher's influential experiences, and construct a preliminary set of questions unique to each teacher. In general, teachers' interviews asked them to describe valuable statistics and variation learning experiences, the content learned during experiences, beliefs about why the experiences benefitted their learning, influential people associated with the experiences, actions taken in response to the experiences, and how the experiences changed the way they thought about statistics or variation. In the second interview (Int II), teachers responded to questions about why the events transpired the way they did.

30.3.2 Data Analysis

Data analysis followed systemic procedures recommended for phenomenological studies (Moustakas, 1994). For teachers who exhibited integrated reasoning about variation, we examined their context interviews, EHCs, and CIs to find evidence of experiences related to developing understanding of variation. We recorded experiences that teachers identified as important for learning about variation and their perceptions of characteristics that helped or hindered development. We sought evidence of elements related to transformative learning such as events that triggered dilemmas, critical reflection, rational discourse with self or others, seeking knowledge related to statistics, experimenting with new roles, and changes in assumptions related to the teaching and learning of statistics. In the first phase of analysis,

interview passages for each teacher were grouped according to reasoning perspectives and elements of transformation. During the next phase, for each perspective, we organized statements into themes and grouped them to produce a factual description of each teacher's phenomenon in relation to the elements of transformative learning. We extracted the essence of the phenomenological experience by reading and rereading passages using the constant comparative method. The final stage of analysis involved integrating descriptions of teachers' experiences for each perspective into a composite description. This synthesized description captures the overall essence of the experience of developing understandings of variation and exhibiting integrated reasoning about variation.

30.4 Results

Many of the learning factors identified by the teachers who exhibited robust understanding of statistical variation also applied to the five additional teachers who exhibited integrated reasoning. These factors, including triggers and dilemmas and personal and environmental factors, transcend perspectives. Brief descriptions of these general factors follow; additional information can be found in Peters (2009). Following this brief overview are descriptions of critical factors associated with teachers' developing understandings for each of the three perspectives, including engagement in rational discourse and participation in key activities.

30.4.1 General Factors Related to Learning and Reasoning about Variation

When preparing to teach statistics and when teaching statistics, each teacher experienced triggers that prompted self-awareness of limitations in their knowledge of statistics. Triggers stimulated dilemmas that were resolved through the creation, enhancement, or transformation of specific knowledge, beliefs, or attitudes. When faced with a dilemma, each teacher viewed the dilemma as a learning opportunity, embraced the opportunity, and reacted to the dilemma by forming and following a plan of action to construct statistical knowledge.

Common to all ten teachers were personal factors that may have influenced their learning. Key among these factors was motivation to encounter and resolve dilemmas, reflection on content, and commitment to their students and to teaching. In addition to personal factors, there were environmental factors that were conducive to their learning. Each teacher cited the importance of a "comfortable" learning environment in which s/he could feel free to ask about content questions as those questions arose. They attributed a sense of community to secondary teachers and statisticians who were active in AP Statistics and described the benefits they attributed to membership in this community, including rational discourse.

30.4.2 Factors Related to Learning/Reasoning about Variation from the Design Perspective

Teachers had few experiences with designing studies or considering designed studies before teaching AP Statistics. Aspects of design repeatedly triggered dilemmas. Common learning factors included engaging in rational discourse while reading and comparing textbook descriptions, actively conducting studies and involving students in designing studies, making sense of problems and activities focused on critical design aspects, and authoring materials.

With the prominence of observational and experimental study design in AP Statistics, every teacher consulted textbooks to deepen their understandings and were confronted with their limited knowledge when preparing to teach the course. By engaging in rational discourse with authors and themselves and reflecting on the premises behind the content, teachers resolved their dilemmas. Dustin, for example, encountered different terminology for similar concepts such as confounding, lurking, and extraneous. “I’d sit there with three or four books open, and I’d read...try to figure out, okay, where the points of commonality were, where they were different” (Int II). He suggested that by critically reading, rereading, and considering textbook passages, divergent images of lurking, confounding, and extraneous transformed into a unified conception of variables different from the independent variable(s) of interest. Cheyenne also indicated how “different ways the authors will write things about blocking” (Int II) allowed her to consider differing views to enhance her understanding. Teachers compared authors’ perspectives with their emerging understandings and considered authors’ arguments as part of resolving their dilemmas to advance their understandings.

Opportunities for teachers and students to design experiments were important for teachers to develop integrated reasoning about variation. Teachers suggested that they generated more ideas about sources of variability and ways to control variability by conducting experiments collaboratively than by reading about similar experiments. For example, Ivy discussed design considerations that arose when conducting an experiment with colleagues: “We shot rubber bands...depending on the angle. What do you notice, does it matter...how many times you shoot the rubber band? Because it loses elasticity” (Int I). Dana also learned from student projects to design and conduct studies. Students’ questions prompted her to engage in rational discourse to consider how and why students might design studies to answer their questions. Active and collaborative engagement with designing studies and participating in rational discourse with colleagues or more knowledgeable others to explore the depth and meaning of design aspects led to deep consideration of potential sources of variation such as rubber band elasticity and methods for controlling variation from sources.

Other experiences that advanced teachers’ understanding of design concepts included responding to AP Statistics problems and considering underlying premises. Problems focused on key concepts that were not well understood by teachers triggered dilemmas that were resolved by engaging in rational discourse with more knowledgeable others and by reflecting on the premises underlying the arguments presented. Dilemmas were triggered when Everett, for example, recognized that his

responses to problems about key concepts such as randomization and blocking would not receive full credit. Resolution occurred from discussing the problems and rubrics with more knowledgeable others and assessing and reflecting on the arguments presented. Similarly, Hudson described how “failure to notice the primary issue being variation” in his response to a question “was concerning” (Int II). His dilemma was resolved through a statistician’s use and discussion of a key example. His engagement with the example and consideration of what the statistician identified as the key aspects of the problem enabled Hudson to see the effects of a particular variable and how removing that potential source of variation would better isolate the treatment’s effects.

Engaging vicariously with students in carefully constructed activities focused on fundamental design concepts and issues prompted learning for teachers. Rational discourse with students and reflection on classroom activities enhanced teachers’ understandings of blocking, stratification, and variation. Hudson, for example, noted the benefits of using the same carefully designed set of data to consider completely randomized, nonoptimal block, and optimal block designs for him and his students to learn about how blocking reduces variation. “The larger variation in that case due to mixing plots...potentially interfered with you being able to detect which variety of tree was actually more productive. Versus when you blocked with the nonoptimal blocking scheme...block the correct way...all you were looking at was the difference in tree varieties” (Int II). This activity and students’ engagement with similar activities to explore the benefits of blocking and stratification were also mentioned by Cheyenne and Gavin as powerful for their learning. Important features were comparisons of different methods such as completely randomized or randomized block designs and use of multiple representations to clearly illustrate *how* blocking or stratification reduced variation.

As leaders, many teachers designed and published tasks and activities. Dana, Everett, and Isaac authored some of the aforementioned activities. Each indicated how authoring required even greater clarity and understanding than teaching because readers could not ask clarification questions. Many publications were collaborations with one or more statisticians, and the statisticians raised issues and questions that triggered dilemmas. Collaborative environments, statisticians’ practical insights and willingness to engage in rational discourse, and teachers’ considerations of the insights offered ideal settings for teacher learning. Everett recalled an illustrative experience: “Having a conversation...for like two hours because we had written something, an initial draft of some activity and [statistician] had some things to say... trying to understand what, what she had a problem with took me a long time” (Int I).

30.4.3 Factors Related to Learning/Reasoning about Variation from the Data-Centric Perspective

Teachers often reasoned about variation from the data-centric perspective in their early learning experiences. Most experiences focused on reasoning about procedures such as calculating standard deviation. Although foundational, teachers felt

that the experiences did not lead to deep understanding. Preparations to teach statistics triggered dilemmas that were resolved through active exploration with data using multiple representations and consideration of the affordances offered by each. Technology use enhanced their pursuits. Teachers' experiences to learn and reason from the data-centric perspective were diverse.

With the exception of Hudson, every teacher described early data-centric experiences as involving "cookbook" procedures. To best describe data, open-ended explorations with real data required more than creating graphical representations or calculating statistics. Cheyenne indicated, "I've played around a lot with real data... fluky things that show up in real data...if you had just looked at the mean and standard deviation, you'd have gotten a totally different picture...mean and standard deviation really isn't telling you a lot" (Int II). Multiple experiences with data and questions about how to best describe data led to recognition of shortcomings associated with limited data representations. Cheyenne further noted, "in courts of law you're always looking for a preponderance of evidence" (Int II). She likened generating multiple representations, calculating multiple summary values, and considering affordances and constraints of each as evidence for understanding and describing data.

Teachers benefited from using graphing calculators and computers to explore data. Active exploration of data, "curiosity," and premise reflection provided opportunities to deepen understandings of variation. Cheyenne noted the ease of creating multiple representations to compare variability within and between distributions. "Looking at distributions...set the scales the same. You can put them on the same page so that you can look at them right next to one another" (Int I). In addition to visualization benefits, Blake noted learning potential from computer output. "Regression analysis...I circled every number and started asking myself what in the heck does this number mean?...sum of the squared residuals. Here's the explained. Here's the total...I was drawing for myself pictures of that and what that meant" (Int II). By considering explained and total variation, he "learned what r^2 was" (Int I). Teachers' individual explorations with data, questions about the meaning of the resulting measures and displays, and reflections on their activities enhanced their reasoning about data and variation.

Teachers' variation-related learning experiences were more diverse when reasoning from the data-centric perspective than design or modeling perspectives. Blake, for example, valued experiences with curricula using a transformational approach to consider problems such as how the value of standard deviation changed if five were added to each data value. Gavin cited the importance of a question for which the rubric identified standard deviation as approximate average deviation from the mean, and Everett and Gavin indicated that the AP Statistics focus on describing and comparing distributions using shape, center, and spread influenced their views of distribution and use of representations. Isaac described benefit from programming computers to compute summary values and construct graphical representations of data and from considering multiple data possibilities for the programs to be useful. Other learning resulted from considerations such as nonparametric statistics and sports data.

30.4.4 Factors Related to Learning/Reasoning about Variation from the Modeling Perspective

As with the data-centric perspective, teachers' early experiences with reasoning about variation from the modeling perspective focused on reasoning about procedural calculations such as standard error. As teachers sought to understand and justify the methods, they encountered dilemmas that were resolved through simulations and pursued new dilemmas by studying content beyond introductory statistics to deepen their understandings.

Active engagement with physical simulations followed by technological simulations was important for teachers to develop meaning of inferential concepts such as sampling distribution. Gavin, for example, identified a simulation to explore the sampling distribution of a sample proportion as important for developing his understanding of sampling distribution. "We did the proportion of brown [candies] and then all went up and plotted it [on a class dotplot]... The intended result that we saw, fairly normal distribution, um, was kind of surprising to me" (Int I). The emerging shape of the distribution piqued his curiosity, and subsequent exploration of the sampling distribution of sample means with a physical and technological simulation of penny ages helped him to realize the relationship between the variation of a population and the variation of sampling distribution. Other teachers, including Blake, Haley, and Hudson, cited similar activities as beneficial for learning. Learning resulted from the combination of physical and technological simulation, the power of visualization, questioning of the processes, and reflection. Further attribution for the power of technology came from applets to explore sampling distributions from populations of different shapes and ideas of chance.

To some extent, simulations substituted for theory. The mathematics to prove many of the generalizations teachers used to make inferences about data required significant theoretical statistics study; simulations provided the means to consider premises in the absence of proof. As Gavin noted, simulation "kind of gives more validity to the theory. Because I've never been taught the theory, I don't know the... mathematical justification" (Int II) behind sampling distribution. Teachers used simulations to consider the premises underlying content and processes for reasoning about variation from the modeling perspective. For example, Dana conducted simulations to examine the conditions needed to use inferential procedures for regression and to explore why t procedures were considered to be fairly robust.

Connecting inference with design by considering inferential methods beyond AP Statistics enhanced teachers' understandings of the methods within AP Statistics. Everett, for example, described benefit for understanding blocking from attending a workshop focused on ANOVA. "[The instructor] presented a matrix/vector model for partitioning the variance in a response variable. We started... with the familiar two-sample t test, but then applied the same procedures to an experiment with three treatments and then to an experiment with blocking. Finally I understood how blocking reduces variability" (CI). Isaac indicated how multiple regression led to recognizing the fundamental role of modeling in statistics: "you're constructing a,

um, a model of the behavior of the real world and trying to account...for the variability that you're observing by appealing to other variables" (Int II). Cheyenne and Everett also ascribed importance to nonparametric methods for deepening understandings of sampling distribution and chance and for enhancing their overall images of statistics.

30.5 Discussion

As the preceding discussion highlights, multiple factors contributed to teachers' learning. Primarily, engagement in rational discourse with self, students, colleagues, statisticians, and textbook authors and active engagement to explore content, processes, and premises by solving problems and exploring data with or without the use of technology were important for teachers to develop understandings of variation and to reason about variation in an integrated manner. Subtleties existed in *how* factors enhanced learning within perspectives.

Within perspectives, rational discourse with others and with oneself differed for developing understandings to reason about variation. To reason from the design perspective, teachers valued rational discourse with more knowledgeable others, such as statisticians, and with other colleagues and students while designing and conducting studies. Benefits came from considering others' perspectives to understand underlying premises related to potential sources of variation from context and error and ways to control that variation. Teachers' reasoning about variation from the data-centric perspective tended to occur while engaging in rational discourse with themselves. They reflected on the meanings attributed to summary measures such as standard deviation and the processes used to describe and compare distributions and justify informal inferences from data. Similarly, rational discourse when reasoning from the modeling perspective tended to occur within teachers even though interactions with others were necessitated when conducting physical simulations. Self-discourse focused on discerning the statistical principles underlying simulated results, and technology provided alternative perspectives of statistical concepts. Unlike reasoning from the data-centric perspective, in which teachers could generate multiple perspectives to provide their own justifications, the technology produced multiple perspectives of sampling distribution for informal justification of inferential methods and for considering chance.

The roles of tasks and activities in teachers' learning revealed similarities and differences. Across perspectives, teachers' learning was enhanced by reasoning about variation when engaged with tasks and activities that attended to fundamental statistical concepts and principles and key aspects of variability. The focus of learning, however, differed when reasoning from different perspectives. Teachers' initial limited experiences with design led them to focus on content and processes when engaging with activities and tasks to construct understanding for new ideas or to enhance minimal understandings. As teachers developed understandings, their learning continued from vicarious engagement with students' activities and authoring materials as their foci shifted to justification and connecting design with data

analysis and inference. Because teachers knew content and processes associated with the data-centric perspective, their learning focused on examining the underlying premises and engaging with the art of statistics to tell the story of data. Hands-on explorations with data and use of technology tools to generate multiple representations led teachers to make and justify informal inferences. Teachers were proficient at procedurally reasoning about variation from the modeling perspective prior to teaching AP Statistics but felt they needed to develop understandings to explain the rationale for the procedures and the logic of inference to better teach statistics. Simulations helped to develop rationale and provide justification for methods such as calculating standard error.

One of the more striking features of teachers' identified learning experiences is that the experiences are accessible to most teachers. Although these ten teachers may more readily embrace learning opportunities than other teachers, this study provides insights into the types of triggers that may stimulate dilemmas and the types of experiences that may enable teachers to resolve dilemmas as a step toward meeting larger curricular and societal goals.

References

- Batanero, C., Burrill, G., & Reading, C. (Eds.). (2011). *Teaching statistics in school mathematics—Challenges for teaching and teacher education*. Dordrecht, The Netherlands: Springer.
- Brookfield, S. (1990). Using critical incidents to explore learners' assumptions. In J. Mezirow et al. (Eds.), *Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning* (pp. 177–193). San Francisco: Jossey-Bass.
- delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4(1), 55–82.
- Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, 4(1), 92–99.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (1998). "Teachers, schools, and academic achievement" (Working Paper No. 6691). New York: National Bureau of Economic Research.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371–406.
- Martyn, K. K., & Belli, R. F. (2002). Retrospective data collection using event history calendars. *Nursing Research*, 51, 270–274.
- Mezirow, J. (2000). Learning to think like an adult: Core concepts of transformation theory. In J. Mezirow et al. (Eds.), *Learning as transformation: Critical perspectives on a theory in progress* (pp. 3–34). San Francisco: Jossey-Bass.
- Moustakas, C. (1994). *Phenomenological research methods*. Thousand Oaks, CA: Sage.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Peters, S. A. (2009). *Developing an understanding of variation: AP Statistics teachers' perceptions and recollections of critical moments* (Doctoral dissertation). University Park: The Pennsylvania State University.
- Peters, S. A. (2011). Robust understanding of statistical variation. *Statistics Education Research Journal*, 10(1), 52–88.
- Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York: Teachers College Press.
- 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.

Chapter 31

Preservice Teachers' Difficulties with Statistical Writing

Min-Sun Park, Mimi Park, Eun-Jung Lee, and Kyeong-Hwa Lee

31.1 Introduction

Statistical literacy has been emphasized as a goal of statistics education. Statistical literacy is the ability to interpret, critically evaluate, and communicate statistical information, arguments, and data (Gal, 2004, p. 70). In information-laden societies, statistical literacy is necessary for citizens since they have to choose and interpret useful information for them. With the emphasis on statistical literacy, the importance of communication is receiving a lot of attention. In addition to the generic reason that it improves statistical literacy, communication is important in statistics for many other reasons. Since statistics is concerned with information about the real world, people have to be able to take problems vaguely conceived in natural language terms through the statistical investigation and analysis cycle to arrive at conclusions that they can successfully communicate to others in natural language (Phillips, 2006).

Communication involves listening and speaking, reading and writing, and representing (Begg, 1997, p. 19). Among these, “writing” is very important, since it is mainly used as a way of assessment (Biehler, 2007; Truran, 1998; Weldon, 2007), and it is also the last step of presenting what people conclude from data analysis. Nevertheless, students usually have trouble with expressing what they understand through writing (Pierce & Roberts, 1998, p. 1202). Many researchers have investigated

M.-S. Park (✉) • M. Park • E.-J. Lee
Graduate School of Seoul National University, Seoul, South Korea
e-mail: minsun26@gmail.com; mimipark1127@gmail.com; ejlee13@snu.ac.k

K.-H. Lee
Seoul National University, Seoul, South Korea
e-mail: khmath@snu.ac.kr

ways to facilitate students' writing (Francis, 2005; Parke, 2008; Peck, 2005). However, researches usually suggest guidelines or particular writing formats, rather than discuss the parts that students cannot express through writing and suggest a solution. Thus, it is necessary to investigate which parts are being excluded from their writing and present a way to bridge the gap between their thinking and writing.

In this chapter, by looking at both written and verbal language, we examined the missing information in preservice teachers' writing even though they understood and considered. For the investigation, a written task about the variance concept was carried out and interviews were conducted. We compared both written and verbal language in the preservice teachers' use of terms, presentation of factors of the concepts that they had considered, and presentation of the solving strategies that they had used. From the results, we suggest instructional ideas for teaching writings and some consideration points for writing assessment.

31.2 Literature Review

In mathematics education, there are some studies that argue that writing can support the problem solving by improving meta-cognitive ability. Pugalee (2004) compared ninth-grade students' written and verbal descriptions of their algebraic problem-solving processes. Through the comparison, he tried to find the connection between problem solving and writing. As a result, students who wrote descriptions of their thinking were significantly more successful in the problem-solving tasks than the students who verbalized their thinking. Differences in students' use of both types of language showed that writing can be an effective tool in supporting meta-cognitive behaviors.

In statistics education, writing is emphasized as well. Lipson and Kokonis (2005) showed that a writing task may be classified as a meta-cognitive activity, and in it of itself provides a means of facilitating the development of conceptual understanding in students. Parke (2008) also investigated the influence of writing. Individual writing assignments, small group activities, and a student-led scoring activity enhanced students' writing as well as their reasoning, understanding, and confidence. Writing tasks encouraged students to take a holistic view of the statistical process (Lipson & Kokonis, 2005, p. 8).

Accordingly, there are many studies presenting ways to facilitate students' writing. Francis (2005) presented an approach that involves giving students a process to follow, clear instructions on the sort of language which is appropriate, and some model reports to use as a guide. Peck (2005) also suggested some ways of facilitating students' writing: being explicit about what is needed for good communication in different settings, emphasizing the importance of context, asking questions that require explanation and interpretation throughout the course, not accepting "mechanics only" answers as correct on homework or exams, encouraging students to read as well as to write, and asking students to write about statistical processes.

Even though both the importance of statistical writing and the way of facilitating students' writing are issued continuously, students still have much trouble with writing. Francis (2005) pointed out some of the reasons, which are as follows: considering statistics as divorced from the real world rather than a source of information about the real world, not knowing what the statistical analysis is about, experiencing difficulty in writing a cohesive report even when understanding a particular analysis, not understanding some of the subtleties of the language, having difficulty with understanding and using statistical terms correctly, and being unsure of what should be included in writing and what should not. All of these difficulties can be connected to the main ideas of statistical literacy. Especially, these are relevant to statistical knowledge and context knowledge, which are knowledge elements in a model of statistical literacy given by Gal (2004, p. 51).

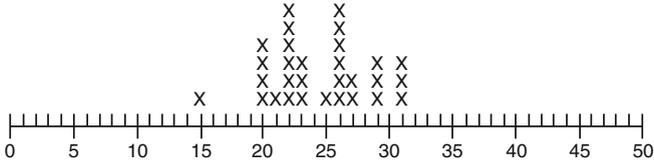
31.3 Methodology

In this research, 44 preservice teachers took a writing assessment and 12 of them, who were selected by the method of stratified sampling, were interviewed. Since preservice teachers should evaluate their students in the future and should be able to respond correctly in the writing assessment and interview, they are appropriate participants.

The task used in the writing assessment was about a variance concept. We were focused on presenting items which require explanation and interpretation (Peck, 2005) to facilitate preservice teachers' writing rather than presenting items which require writing about a simple definition of variance. As a result, there were items asking about the meaning of variance in a particular context, comparing the degree of variances with reasons, and estimating the change of variance when data sets were changed with reasons. The items were taken from previous studies on variability (Canada, 2004; delMas & Liu, 2005; Lee & Meletiou-Mavrotheris, 2003; Watson, Kelly, Callingham, & Shaughnessy, 2003 and CAOS test), and after the pilot study, the items were modified for preservice teachers' understanding. Several items used in the writing assessment are given in Fig. 31.1.

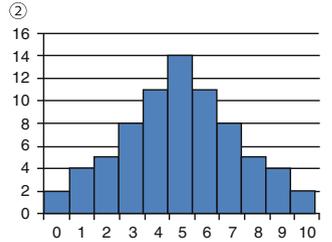
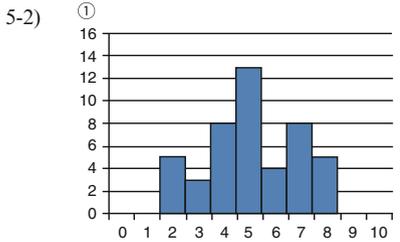
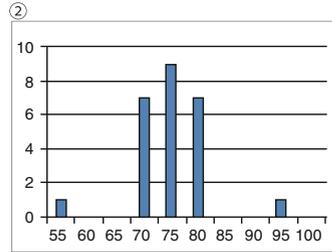
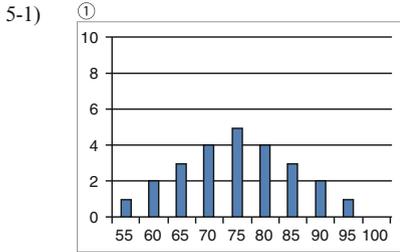
The writing assessment took 40–70 min and the interview took 30–60 min per person. In the interview, preservice teachers were asked about their way of thinking when they took the writing assessment. By analyzing both the written responses and interview results, we tried to find the missing information in their written language compared to the verbal language which revealed what the preservice teachers knew or had considered. There is some possibility that the researcher's reactions during the interview could affect the interviewee's response; also, preservice teachers were able to change their answers after they thought about the question again meta-cognitively. Because of these limitations, we excluded the parts for which preservice teachers changed their answers and spoke retrospectively. The interview was in a semi-structured format, and every interview was recorded and transcribed for analysis.

2. There is a spinning disk that is half-white and half-black. A class did 30 sets of 50 spins and the results for the number of times it landed on the black part are recorded below.

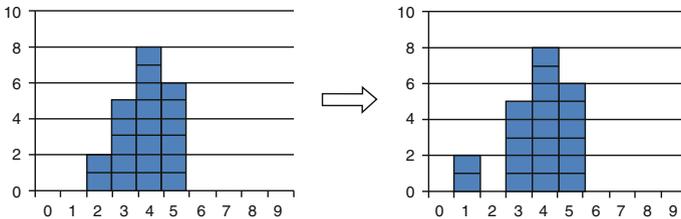


2-1) If you got a variance for the graph above, what does it mean?

5. For each pair of graphs, determine which graph has the bigger variance. Explain your answer.



8. In the following graph, if the two lowest data points had a value of 1 instead of 2, how would the variance be changed? Explain your answer.



11. Of the followings, determine which sentence has the bigger variance. Explain your answer.

- ① Age of trees in a national forest
- ② Diameter of new tires coming off one production line
- ③ Scores on an aptitude test given to a large number of job applicants
- ④ Daily rainfall
- ⑤ Weight of a box of cereal

Fig. 31.1 Examples of several items in written assessment

From the written response and interview transcripts, we compared the following: The subjects' usage of terms; presentation of factors of the concepts that they considered; and presentation of the problem-solving strategies that they used.

31.4 Results and Discussion

After comparing the preservice teachers' written response and interview results, there were the following differences between written language and verbal language: connecting of the terms contextually and conceptually, presentation of various factors of the concepts that they considered, and presentation of the problem-solving strategies that they used. These were not apparently exposed in written language, which means that these are the points with which preservice teachers have difficulty.

31.4.1 *Difficulty in Connecting Terms Contextually*

In the writing assessment, almost all of the items were based on some context. Thus, the preservice teachers were required to interpret the meaning of the variance of the given data sets rather than merely provide the formal definition of the variance. This requirement ensured that the preservice teachers could connect their usage of terms to the context. Several preservice teachers, in their written responses, described the formal definition of variance instead of reflecting the context of the problem. However, in the interview, they showed much understanding of the context (Table 31.1).

Both S7 and S10 described the formal definition of variance in written language. So their terms that were used to show the meaning of the concept did not reflect the context well. In contrast, the preservice teachers considered the mean of the given data sets and focused on the outlier in the interview, which means that they used

Table 31.1 Examples of difficulty in connecting terms contextually

Written language	Verbal language
S7: The degree of the distance of the data values from the mean (Q2-1)	S7: Maybe I was thinking about the mean value that I calculated. I used some process of elimination. I can eliminate the same number of each data which are at the same distance from 25. It was not 25 exactly. Maybe about 24.5 or 25.5, if I remember it correctly... So I think that the variance would be the degree of distance from that point.
S10: The degree of the spread in the result (Q2-1)	S10: In the experiment, somewhat regular values should be given. So 15 seems awkward. If this graph was not given and only the variance value was given, we can calculate the probability without looking at this graph. If the variance seems unusually big, then we can expect that in the experiment, there were some extreme points, like 15. That's what we can know from the variance.

appropriate words connected to the context. As Cobb and Moore (1997) said, data in statistics are not just numbers, but numbers with a context. Therefore, considering the context is essential. Peck (2005) mentioned that the importance of context is a primary reason that communication is such an important aspect of statistics problems. From the result, we examined the preservice teachers’ difficulty in connecting terms contextually in written language.

31.4.2 *Difficulty in Connecting Terms Conceptually*

When communicating about statistics, the usage of terms showing the understanding of the concept which is included in the problem was required. The task was about a variance concept. To show conceptual understanding of the variance, the preservice teachers should properly connect explanatory terms like mean, deviance, and frequency to common words. If they use their own informal terms, then their response would be considered to be the opposite of a conceptual response. Some preservice teachers used their own terms in the writing assessment, even though they connected terms conceptually well in the interview (Table 31.2).

We can see that S9 used his own terms like “pointed” and “thick tails.” He said that he used those words because they were the words that he usually used. However, in the interview, he tried to approach the variance conceptually by connecting terms like mean, data values, and deviance. Likewise, S1 referred to a big range as a “wide spectrum.” However, in the interview, she showed her understanding of the variance by using terms like mean and deviance. Using statistical terms is important since it is an element of the statistical knowledge base in a model of statistical literacy. Moreover, preservice teachers should connect those terms to the common words conceptually to show their understanding of the concept. From Table 31.2, we could see that preservice teachers feel difficulty in connecting terms conceptually.

Table 31.2 Examples of difficulty in connecting terms conceptually

Written language	Verbal language
S9: ① less pointed (Q5-1)	S9: ... In graph ②, the data values are very crowded around 75 and in graph ①, the data values are evenly spread from 75. So I thought ②’s variance is smaller because it is much gathered. I mentioned “pointed” in my reason because that is the term that I usually use.
S9: ① thick tails (Q5-2)	S9: Suppose that the mean is 5. If the values of both ends are big, then the deviance would be big accordingly. Since the values show big deviance, meaning that the values on the very end are big, or have high values, the variance is bigger. That’s why I wrote “thick tails.”
S1: ④ Wide spectrum from no rain to very much rain (Q11)	S1: There are days when there is no rain and there are days when there is very much rain. We have to find the mean of rainfall from those days. So if the mean lies between these days, then the deviance would be big.

31.4.3 *Difficulty in Presenting Various Factors of a Concept*

When dealing with the variance concept, people should consider various factors of variance: mean, data values, frequency, and variability, which can be seen from the distribution. Especially, when trying to compare the degree of variance, they should consider more than two factors. The preservice teachers presented one or two factors related to the meaning of variance in the written assessment; however, in the interview, they explained variance by including various factors that they had actually considered (Table 31.3).

In the writing assessment, S5 only considered the mean and data values. However, from the interview, we could find that he was considering the same value, symmetric graph, and the same range, which are rather various and specialized. Also, S11, in his written response, wrote only about his consideration of the movement of data values. That is a different result from that of the interview in which he mentioned the movement of the mean and the change of the deviance of other data values. He said that he had somewhat considered those factors when he solved the problem. Distribution itself is a multifaceted concept (Bakker & Gravemeijer, 2004). When considering the variability of the distribution, we should integrate various factors of the distribution. Regarding this point, we could find that the preservice teachers had difficulty in writing about integrating the various factors of the concept.

Table 31.3 Examples of difficulty in presenting various factors of a concept

Written language	Verbal language
S5: ① The number of data values and degree from the mean are bigger in ① than ②. (Q5-1)	S5: ... In graph ①, since the size of the data sets are the same, the degree of spread would affect the variance. Obviously, the mean of both graphs is 75, and they are symmetric. If in graph ②, the values of 55 and 95 were 45 and 105, then there would be many things to consider. However, they are 55 and 95 and both the number and data are the same as those of ①. Thus, they do not affect the variance very much ...
S11: The variance gets bigger since the gathered data values are scattered. (Q8)	S11: ... If we look at the shape of the graph, only these data values are moved in this way. Then the mean would be moved in the same way and the distance between these values and the mean would be bigger. If we square those numbers, that is, if we consider the formula of variance, then the value of $(x-m)^2$ gets bigger. So the variance gets bigger.
	R: The explanation you just gave and the writing... When you solved this problem, did you consider what you just said?
	S11: I thought about the movement of the mean. I don't think I considered all of the things specifically. I just thought that the mean would be changed and the deviance would be bigger.

Table 31.4 Examples of difficulty in presenting problem-solving strategies

Written language	Verbal language
S11: ① Values are rather spread to the outer side. (Q5-1)	S11: The end points, which are 55 and 99, are 1 each, and the middle area is empty. In graph ②, all of the values are stacked in the middle. So if we consider the possibility of changing ② to ①, that is, keeping in mind the changing situation, then we can imagine cutting these points in the middle and sending them to the side. We can think like that. R: So you made a transformation? S11: Yes, I changed the data... the shape of the graph. Then now we can think that if the mean is 75, we sent data values around the mean to the side, and the variance got bigger.
S6: ① The frequencies of the values which are far away from the mean are relatively bigger. (Q5-1)	S6: ... If we draw the shape of the graph, ② is much thinner and ① has wide bell-shaped. In my mind, the variance and standard deviation of the wide bell-shaped graph are always big. So I could make a guess. R: How did you make those images? S6: In high school, I learned a lot about normal distribution and I usually draw the graph in a bell-shape. Also, I saw that thinner graphs have small standard deviation and wide graphs have big standard deviation. I applied those images to this problem.
S7: ① The number of values that are far away are bigger than those that are near. (Q5-2)	S7: If we change the order of the data values like a step function, ... If we change these two values, then it would be in the shape of steps. Also, if we send this value to the end, then it would be in the shape of steps. Oh, not steps, but mountains. No, maybe a pyramid? Anyway, it would be in the shape of a mountain. Then both graphs are shaped like a mountain, which is the shape that has the smallest variance. So, intuitively, the first one was much more twisted than the mountain shape, which means that it has a bigger variance.

31.4.4 *Difficulty in Presenting Problem-Solving Strategies*

When comparing variance, people usually use some strategies. delMas and Liu (2005) presented various strategies that students used in comparing standard deviation. The preservice teachers did not mention the strategies that they used in their written responses. However, after the interview, we could find that they indeed used some strategies like changing one graph to the other graph or imagining a bell-shaped graph with which they had already dealt (Table 31.4).

S11, in the writing task, compared the variance of the graphs using the degree of the spread. However, in the interview, he mentioned the strategy that he had used, which was sending some of the data values in one graph to apply them to the other graph. By checking the movement of the data values, he could compare the variance of the graphs. S6 also presented some factors of variance in writing, which differed from the interview result where she mentioned her image of bell-shaped graphs. In the case of S7, he mentioned the strategy of transforming the graph like S11 did.

Also, by saying that the “mountain shape is the one that has the smallest variance,” he also used some images in his mind like S6 did. All three preservice teachers presented strategies and they were focusing on the informal aspect of the distribution that Bakker and Gravemeijer (2004) emphasized. However, these considerations were not communicated well in the writing assessment, which means that the preservice teachers have difficulty in presenting problem-solving strategies.

31.5 Conclusion

All four aspects that are presented in the results section are important aspects that students should have for statistical literacy. Students should be able to connect terms contextually and conceptually, and present various factors of a concept and the problem-solving strategies that they considered or used. From the interviews, we could find that the preservice teachers considered all of the four aspects; however, they could not explain them in writing very well though they were able to discuss the aspects during the interviews. If we compare those difficulties with the result of Francis (2005), difficulty in connecting terms contextually is relevant to “considering statistics as divorced from the real world rather than a source of information about the real world,” and difficulty in presenting various factors of a concept and the problem-solving strategies that they considered or used is relevant to “being unaware of what belongs in writing and what does not.” Francis also mentioned the difficulty of understanding and using statistical terms correctly; however, in this chapter, we found the additional difficulty of connecting statistical terms and common words conceptually rather than the direct use of statistical terms.

This study included tasks which require explanation and interpretation to facilitate preservice teachers' writing (Peck, 2005). However, the gap existing between the preservice teachers' written and verbal language indicates that they still have difficulty in writing. Further research should consider other ways of facilitating students' writing. The difficulties of presenting various factors of a concept and the problem-solving strategies that they considered or used could be resolved if we include some items that ask students to write about statistical processes, as Peck mentioned.

References

- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer Academic.
- Begg, A. (1997). Some emerging influences underpinning assessment in statistics. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 17–25). Amsterdam, The Netherlands: IOS Press.

- Biehler, R. (2007). Assessing students' statistical competence by means of written reports and project work. In *Proceedings of the IASE Satellite Conference on Assessing Student Learning in Statistics*. Guimaraes, Portugal, August 2007.
- Canada, D. (2004). *Elementary Preservice Teachers' Conceptions of Variation*. Unpublished doctoral dissertation. Portland State University, Portland.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematics Monthly*, *104*, 801–823.
- delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, *4*(1), 55–82.
- Francis, G. (2005). An approach to report writing in statistics courses. In *Proceedings of the IASE Satellite Conference on Statistics Education and the Communication of Statistics*. Sydney, Australia, April 2005.
- Gal, I. (2004). Statistical Literacy. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 47–78). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Lee, C., & Meletiou-Mavrotheris, M. (2003). Some difficulties of learning histograms in introductory statistics. In 2003 *Proceedings of the American Statistical Association, Statistics Education Section* [CD-ROM], (pp. 2326–2333). Alexandria, VA: American Statistical Association.
- Lipson, K., & Kokonis, S. (2005). The implications of introducing report writing into an introductory statistics subject. In *Proceedings of the IASE Satellite Conference on Statistics Education and the Communication of Statistics*. Sydney, Australia, April 2005.
- Parke, C. (2008). Reasoning and communicating in the language of statistics. *Journal of Statistics Education*, *16*(1). Retrieved from www.amstat.org/publications/jse/v16n1/parke.html.
- Peck, R. (2005). There's more to statistics than computation—Teaching students how to communicate statistical results. In *Proceedings of the IASE Satellite Conference on Statistics Education and the Communication of Statistics*. Sydney, Australia, April 2005.
- Phillips, B. (2006). Statistics education and the communication of statistics: A report on the IASE/ISI satellite meeting, 2005. In *Proceedings of the 7th International Conference on Teaching Statistics*. Salvador, Bahia, Brazil, 2006.
- Pierce, R., & Roberts, L. (1998). Introductory statistics: Critical evaluation and clear communication. In *Proceedings of the 5th International Conference on Teaching Statistics*. Singapore, June 1998.
- Pugalee, D. (2004). A comparison of verbal and written descriptions of students' problem solving processes. *Educational Studies in Mathematics*, *55*, 27–47.
- Truran, J. (1998). Using students' writings to assess their cognitive and affective development in an elementary economic statistics course. In *Proceedings of the 5th International Conference on Teaching Statistics*. Singapore, June 1998.
- Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, *34*, 1–29.
- Weldon, L. (2007). Assessment of a writing course in statistics. In *Proceedings of the IASE Satellite Conference on Assessing Student Learning in Statistics*. Guimaraes, Portugal, August 2007.

Chapter 32

Teachers' Questions in the Statistics Class

Lucia Zapata-Cardona and Pedro Rocha-Salamanca

32.1 Introduction

There is empirical evidence that teachers' knowledge strongly influence students' learning (Hill & Ball, 2004). Teachers' knowledge has been a matter of interest in research. Some authors have studied teachers' knowledge as the result of teacher preparation and teacher experience while others have studied what teachers do with that content knowledge. This is the abilities of the teacher to understand and use subject matter knowledge to accomplish the tasks of teaching. In other words, how the content knowledge is used in the class. In this view, teachers' knowledge goes beyond subject courses taken in college or results in content tests. Consequently, teachers' knowledge cannot only be pictured by means of teachers' subject matter skills but by means of teachers' use of specific representations, explanations, and analysis of students' solutions. The skills the teacher places on the design of student' assignment, management of class discussion, and questioning in the class are better indicators of teacher knowledge than that offered by the teachers' subject matter skills.

In this research, we were interested in the teachers' knowledge that is displayed in the statistic class at different educational levels. We studied different aspects that account for statistical knowledge for teaching such as teachers' explanations and representations, type of examples displayed, ways of dealing with

L. Zapata-Cardona (✉)
Universidad de Antioquia, Antioquia, Colombia
e-mail: luzapata@ayura.udea.edu.co

P. Rocha-Salamanca
Universidad Distrital Francisco José de Caldas, Bogotá, Colombia
e-mail: pgrocha@udistrital.edu.co

students' difficulties, class design and evaluation, and type of questions exhibited. In spite of the different aspects studied, in this chapter and because of space constraints, we will only report the results related to the teachers' questions asked in the statistics class.

32.2 Brief Literature Review

Questions in the classroom play a very important role in the teaching and learning. Effective questioning reveals the teachers' knowledge about the subject and about students' information processes (Moyer & Milewicz, 2002). The importance of teachers' questions in the statistics classroom cannot be underestimated. The questions the statistics teacher exhibits in class are a very good way to look at the class quality and at the teachers' knowledge. These questions are important in shaping the classroom atmosphere that allows the development of students' statistical thinking. Questions in the statistics classroom have different intentions. Some questions can be used for provoking students thinking, for analyzing students' thoughts, for initiating students' discussions, and for reviewing material. Several studies have focused on teachers questioning in the classroom. Some scholars have categorized and count the teachers' questions in an effort to predict students' achievement (Adedoyin, 2010; Cotton, 1989; Hancock, 1995). Others have focused on the different types of questions and have studied if different kinds of questions lead to different levels of students thinking. Some of these studies have focused on high-order questions (Brualdi, 1998), factual questions (Vacc, 1993), open-ended questions (Hancock, 1995; Vacc, 1993), and probing questions (Newmann, 1988).

Research on teachers' questioning in the classroom has spread out in different subject matters and has been linked to the teaching process. The majority of the studies in this topic have been carried out mainly by observing teachers in the classroom (Brock, 1986). However, there are also studies that have used indirect methods (Adedoyin, 2010; Çakmak, 2009). One study, for example, developed a survey for preservice teachers in which they were asked the reasons for questioning in the classroom. Findings revealed that preservice teachers think that questions are important *to control whether pupils understand or not* and *to make teaching active* but are less important *to manage the classroom*. In this study, only 6.19 % of the participants considered that questioning in the classroom is important to *encourage students to think* (Çakmak, 2009).

In fields like science education, researchers have studied the types of questions displayed by the teacher according to the required knowledge to answer the questions (Chi, de Leeuw, Chiu, & LaVancher, 1994). The researchers found that teachers ask different kind of questions in the classroom. Some of them are *literal* that only require recall of facts, others require *inference of understanding*, others *inference of knowledge*, and only few of them require that the students use their knowledge in a *practical application*.

In language education, teachers' questions have also been a matter of concern. A study in English as a Second Language classrooms showed that there is substantial difference in the learning when students are asked *display questions* (answers already known by the students but calling for the recognition or recall of factual information) and when they are asked *referential questions* (suggesting evaluation or judgment) (Brock, 1986). Such a study showed a strong positive relationship between the cognitive level of the teachers' question and the cognitive level, length, and syntactic complexity of the students' response.

A study in mathematics education focused on the students' perceptions in relation to teachers' questions and students' achievement in mathematics. The study applied a Likert scale to 471 junior secondary students from Botswana. Results revealed that teachers' questions have no effect on the students' mathematics performance. This might indicate that the learning of mathematics in junior secondary schools is not promoted through teachers' questions in the classroom (Adedoyin, 2010). Although the study was focused on classrooms questions, it did not have data from the real classrooms but from students' perceptions.

As the brief literature review shows, teacher questioning in the classroom is a problem that has been studied in the context of different subject matters and the results have shown that the teachers' questions can be considered as the most powerful device to lead, extend, and control communication in the classroom. When used well in teaching, questions function to activate thinking. By using questioning and other appropriate teaching strategies teachers can facilitate the development of critical thinking, decision making, and problem solving in students. In spite of the results from research and the curriculum reforms, teachers still ask similar questions in the classroom.

There is a recent reform in mathematics curriculum in Colombia. We have gone from a mathematics curriculum that emphasized the numerical component to an integrated curriculum that focus evenly on arithmetic, algebra, geometry, statistics, and measurement (MEN, 2006). Particularly, the statistical component of the new curriculum emphasizes more in the skills of interpreting, reasoning, predicting, comparing, justifying, and inferring than in applying algorithms and having factual knowledge. The new demands of the Colombian curriculum suggest new ways of teaching. Focusing on the teachers' questions in the statistics classrooms is crucial at this historical moment in Colombia since it would give us a flavor of how the curricular reform is being adopted in the statistics classroom and it would uncover the statistical knowledge of the teachers.

32.3 Methodology

We gathered information from: Eighteen Colombian elementary through high school statistics classrooms observations, teachers' interviews before and after a statistics class (eight male and ten female), teachers' artifacts, and discussions in learning communities. Classroom observations were kept in video and interviews

were kept in audio; classes and interviews were transcribed *verbatim*. The main source of data for this particular report was classroom observations. We analyzed the information with the aim of Atlas.ti software. Each researcher, independently, saw the video of each class and, with the help of the transcripts, the episodes where teachers asked questions were coded. Then the researchers got together to compare the coding and there was agreement in most of the codes. Those where there was disagreement were discussed up to finding a common ground. Once the coding was done, the research team got together in the learning community to study each of the questions asked by the teachers and to construct the emerging categories (as it is suggested by Hernandez-Sampieri, Fernandez-Collado, & Baptista-Lucio, 2008). The learning community involved the principal investigator, coinvestigators, research assistants, and some of the in-service statistics teachers who participated in the study. The teachers' questions were classified in four emerging categories according to the purpose of the question and the level of knowledge required for the students to answer them.

32.4 Results and Discussion

A total of 267 questions asked by the eighteen in-service statistics teachers were identified. The coding, organization, and analysis of the questions that took place in the classrooms seem to be indicative of a pattern. The questions led to four emerging categories that are explained as follows:

Close Questions: These are the questions the teacher asked when interested in getting a specific answer associated with knowledge of facts or with short answers that did not require further elaboration. Most of the time these questions were related to expressions such as: how many, which one, what is. From time to time teachers specifically asked for definition of concepts. Some examples of questions in this category are: How many red cards are there in a deck of cards? What is a sample space? What is probability for you? How can we define random? What is the smaller value I can get in a probability calculation?

Procedural Question: Sometimes teachers were interested in examining the way of carrying out certain algorithmic routine. Examples of these questions are: How can you find the average? How can we calculate probability? How can we get a specific value of the sample space?

Monitoring Questions: Sometimes teachers used these questions to check if students were following certain explanations. These questions are intended for examining students' pace in the class but not necessarily to study students' understanding in a deep way. *Monitoring Questions* might resemble *Close Questions*. A central characteristic to differentiate them is to consider that the purpose of *Monitoring Questions* is to check if students are following the class and those questions need to be related to the current class discourse. Some examples of these questions are: What is the sample in

this example? What is what we need to find in this exercise? Does somebody have any question up to hear? What is the population we are talking about here?

Analysis Questions: These are high-order questions and different versions of them included teachers demand for justification, reasoning, prediction, or decision making. Teachers explored the reasons students give for certain actions or decisions, checked on students' ability to use information for making conclusions, pushed students to reason about the validity of certain information, gave information, and asked students to make decisions. Although these types of questions were very absent in the classroom observations, we present some examples: What are the reasons you have to say that player A has some advantages over player B? What would be the usefulness of knowing the probability of an event? What do you think of your classmate reasoning? We toss a die, if we get a divisor of three I clean the dishes, but if we get a divisor of two you clean the dishes. Is this proposal fair?

Table 32.1 shows the percentage of questions of each type asked for the teachers. The percentage of questions varied from teacher to teacher but there seems to be a pattern where teachers privilege *Close*, *Procedural*, and *Monitoring Questions* over *Analysis Questions*. Only one teacher had an even pattern among the questions but the total percentage of all questions reveals that there is a strong privilege for *Close Questions*. The table also indicates that very few teachers asked students *Analysis Questions*.

The classification revealed that *Close Questions* took place in the classrooms 52.1 % of the times, *Procedural Questions* 15.4 %, *Monitoring Questions* 19.1 %, and *Analysis Questions* 13.5 %. These results make public that the statistics class strongly privilege low-order thinking questions while higher order thinking questions that stimulate students' statistical reasoning are sporadic. Additionally, these results make clear that the statistics class in Colombia is short to be considered in agreement with the last curriculum reform that claims for the development of students' statistical reasoning more than the learning of facts and procedures.

The results from this study can be explained by taking different approaches. First, low-level thinking questions take little time to prepare while high-level thinking questions are demanding and require a well-integrated teachers' knowledge about the subject, the students, and the class management. Second, teachers might not be aware of the level of thinking promoted through their questions. Most of the participant teachers in the study were surprised when the results were shared. The teachers concluded that they might not be challenging their students enough. Third, statistics teachers might not feel comfortable with the topic taught and they might prefer to ask questions in which they have full control and do not risk with unexpected answers. A recent study revealed that 20 % of statistics teachers in Colombia have never taken a single course in statistics and 50 % only have taken one course (Zapata-Cardona & Rocha-Salamanca, 2012). The abundant interaction in which the teacher has the control of the class with *close questions* is also a revelation of the teacher knowledge. The amount of *close*, *procedural*, and *monitoring questions* in the classroom could reveal the teachers' lack of confidence in the subject matter they teach. Open-ended questions were scarce perhaps due to the unpredictability of

Table 32.1 Percentage of questions of each type asked by the statistics teachers

Type of questions						
Teacher	Grade	Class topic	Close	Procedural	Monitoring	Analysis
Carlos	Sixth	Graphic representation of data	55.6	11.1	33.3	0.0
William	Ninth	Probability	91.7	8.3	0.0	0.0
Mosquera	Eighth	Graphic representation of data	12.5	12.5	75.0	0.0
Susana	Ninth	Counting techniques	12.5	50.0	25.0	12.5
Marta	Tenth	Spread measurements	50.0	0.0	16.7	33.3
Rodrigo	Tenth	Central tendency measurements	0.0	0.0	0.0	0.0
Gloria	Tenth	Probability	50.0	25.0	16.7	8.3
Claudia	Eighth	Statistical concepts	25.0	25.0	25.0	25.0
Pablo	Eleventh	Probability	44.4	25.9	11.1	18.5
Fredy	Tenth	Probability	45.5	18.2	0.0	36.4
Diana	Ninth	Central tendency measurements	40.5	14.3	21.4	23.8
Oswaldo	Tenth	Graphic representation of data	39.1	4.3	47.8	8.7
Ricardo	Seventh	Central tendency measurements	83.3	16.7	0.0	0.0
Rosalba	Fifth	Graphic representation of data	71.4	0.0	28.6	0.0
Marcela	Seventh	Probability and graphic representation	63.6	13.6	9.1	13.6
Zoraida	Third	Data collection	88.9	0.0	11.1	0.0
Carmen	Fourth	Data collection	100.0	0.0	0.0	0.0
Sonia	Tenth	Probability	55.2	24.1	13.8	6.9
<i>Total percentage</i>			52.1	15.4	19.1	13.5

Teachers' names are pseudonyms to protect the confidentiality of the participants

the students' responses. Finally, all the participant teachers in this study were aware of the curriculum reform and what they did in class was supposed to follow the curriculum requirement; however, the type of questions they encouraged in the classroom were far to be considered promoters of the statistical reasoning claimed in the reform. Perhaps what they do in the class reflects how they interpret the curriculum reform. This could indicate that a successful adoption of a reform requires a strong participation of reflective teachers not only as consumers but as creators.

32.5 Implications

Results from this study suggest that teacher reflection is needed. Teachers might not be aware of the level of thinking promoted in their classroom by means of their questioning if somebody else does not help them see the implications of the questions. Teachers can sharpen their questioning skills by becoming familiar with different typologies of questions to help students think critically. Teachers should be provided with reflective training in developing their questioning techniques. There is evidence that with training teachers can modify their questioning skills so that they can ask high-order questions in the classroom (Brock, 1986). A good question is a powerful teaching tool and teachers should know how to use questions to teach effectively. Good questioning requires technical knowledge.

These results also suggest that Colombian in-service statistics teachers use questions in the classroom for many different reasons but *encourage students to think* is one of the less valuable reasons. The majority of the questions proposed by the statistics teachers in the classroom focused on low-level thinking which just check for students' knowledge of facts or procedures. The low frequency of high-level thinking questions could be an indicator of the limitation of the teachers' knowledge. Students should be challenged with questions that are not only checking knowledge but offer opportunities to go beyond that. Those opportunities should promote the development of the statistical reasoning as it is required in the recent curriculum reform in Colombia. However, although good questioning techniques are favorable skills in teachers, developing students' statistical reasoning requires well-developed and well-integrated teachers' subject matter knowledge. Perhaps professional development programs for teachers are even more complex than anticipated and they should be focused on holistic training and not only on isolated aspects of teaching.

Acknowledgment This research was supported under a grant from COLCIENCIAS—Colombian Institute for Science and Technology Development “Francisco José de Caldas”—contract 782 of 2009 Code 1115-489-25309.

References

- Adedoyin, O. (2010). An investigation of the effects of teachers' classroom questions on the achievements of students in mathematics: Case study of Botswana community junior secondary schools. *European Journal of Educational Studies*, 2(3), 313–329.
- Brock, C. A. (1986). The effects of referential questions on ESL classroom discourse. *TESOL Quarterly*, 20(1), 47–59.
- Brualdi, A. (1998). Classroom questions. *Practical Assessment, Research and Evaluation. ERIC/AE Digest*, ED422407, 6(6).
- Çakmak, M. (2009). Pre-service teachers' thoughts about teachers' questions in effective teaching process. *Elementary Education Online*, 8(3), 666–675.
- Chi, M., de Leeuw, N., Chiu, M., & LaVanher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 437–477.

- Cotton, K. (1989). *Classroom questioning*. School Improvement Research Series (SIRS). Portland, OR: Northwest Regional Educational Laboratory.
- Hancock, C. L. (1995). Enhancing, mathematics learning with open-ended questions. *The Mathematics Teacher*, 88(6), 496–499.
- Hernandez-Sampieri, R., Fernandez-Collado, C., & Baptista-Lucio, P. (2008). *Metodología de la investigación*. México: McGrawHill.
- Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's Mathematics Professional Development Institutes. *Journal for Research in Mathematics Education*, 35(5), 330–351.
- MEN. (2006). *Estándares básicos de competencias en lenguaje, matemáticas, ciencias y ciudadanas*. Santa Fe de Bogotá, Ministerio de Educación Nacional: Revolución Educativa Colombia Aprende.
- Moyer, P. S., & Milewicz, E. (2002). Learning to question: Categories of questioning used by preservice teachers during diagnostic mathematics interviews. *Journal of Mathematics Teacher Education*, 5, 293–315.
- Newmann, F. (1988). A test of higher-order thinking in social studies: Persuasive writing on constitutional issues using NAEP approach. *Social Education*, 54(4), 369–373.
- Vacc, N. (1993). Implementing the professional standards for teaching mathematics: Questioning in the mathematics classroom. *Arithmetic Teacher*, 4(2), 88–91.
- Zapata-Cardona, L., & Rocha-Salamanca, P. (2012). *Qué es y qué debe ser en Educación Estadística [What is it and what should it be in Statistics Education]*. Bogotá, Colombia: Research report supported under a grant from COLCIENCIAS.

Chapter 33

Statistical Training of Pre-service Teachers with Application in School Practice

Miriam Utsumi, Irene Cazorla, and Verônica Yumi Kataoka

33.1 Introduction

In Brazil, teachers are expected to have an understanding of probability and statistics, as suggested by the National Curricular Parameters for correct teaching. Those concepts are in section known as “Treatment of Information” (data handling) in elementary schools (Brasil, 1997) and middle schools (Brasil, 1998), and in the “Data Analysis” section for the high school (Brasil, 2002, 2006). However, results for the last National Examination of Performance to Students from Mathematics Courses show that the performance of pre-service teachers (undergraduate students) is much lower than expected, with the further problem that these graduates are legally authorised to work with students in elementary and high schools without having sufficient knowledge, especially relating to the didactical aspects, to provide a quality education.

With the aim of reversing this situation, in 2007 the Brazilian government created the Scholarship Program for Pre-service Teachers (PIBID) in order to support and enhance the learning of teaching by pre-service teachers, by placing them in

An extended version of this chapter was published in: Utsumi, M. C., Cazorla, I. M., & Kataoka, V. Y. (2014). Statistical training of pre-service teachers with application in school practice. *American Review of Mathematics and Statistics*, 2(1), 55–66.

M. Utsumi (✉)

Universidade de São Paulo, São Carlos, Brazil
e-mail: mutsumi@icmc.usp.br

I. Cazorla

Universidade Estadual de Santa Cruz, Ilhéus, Brazil
e-mail: icazorla@uol.com.br

V.Y. Kataoka

Universidade Estadual de Santa Cruz, Ilhéus, Brazil
e-mail: veronicayumi@terra.com.br

schools, with responsibility for preparing and implementing activities in partnership with adviser professors at the university and supervisor teachers at school.

Several researchers have reported experiences with pre-service teacher training and studies have been developed about these experiences. For instance, Groth and Bergner (2004) reported on teaching statistical sample, Canada (2006) about variation in a probability context and Leavy (2010) about informal inferential reasoning.

In this chapter, we present a discussion about an application of probability and statistics activities by a group of four pre-service teachers in mathematics, all participants in PIBID with students from the sixth to ninth years of schooling (from 11 to 14 years old) at two public schools in the State of São Paulo.

33.2 Literature Review

As in other countries, statistics and probability are part of the mathematics curriculum in Brazilian schools. According to Batanero, Godino, and Roa (2004) this is due to the usefulness of statistics and probability in daily life, its instrumental role in other disciplines, the need for basic stochastic knowledge in many professions and its role in developing critical reasoning.

Despite this, Shaughnessy (1992) points out that the unfamiliarity of mathematics teachers during their training with statistics is a major obstacle to successful teaching and school learning of this subject. Indeed, Cazorla (2006), Contreras, Batanero, Díaz, and Fernandes (2011), Nicholason, Road, and Darnton (2003), Peck and Gould (2005) all argue that teachers from mathematics graduate courses, sometimes have some basic training in probability and statistics, but generally are not trained in issues related to teaching these subjects. According to Viali (2008), most of major mathematics courses in Brazil offer only a single 60–75 h course on descriptive statistics and probability, and this rarely deals with aspects of teaching statistics. For example, in the “Introduction to the Theory of Probability” course, the goals involve familiarising the student with probabilistic reasoning and providing a basic knowledge for the proper understanding of statistical methods.

In this study, the statistics course provided is the Theory of Probability at a higher education level without discussing about didactical aspects of statistical thinking. This way, it will be difficult for prospective teachers to teach in basic education as shown below. An analysis of the contents of the course (frequency and probability, conditional probability and independence, random variables, discrete distributions (Uniform, Bernoulli, Binomial, Geometric, Hypergeometric and Poisson), continuous distributions (Uniform, Exponential, Gamma, Normal and *t*-Student), Normal approximation to the Binomial, *n*-dimensional variables, hope, variance, covariance, Markov inequality and Central Limit Theorem)—suggests that there are many topics covered, but the topics refer only to probability and statistics at an advanced level, and therefore do not take account of the level of content that these teachers will have to teach in middle school, much less how to teach such content.

Table 33.1 Statistic and probability content in the curriculum for Brazilian schools

Content	Sixth to seventh grade	Eighth to ninth grade
Table and graphs	To read and understand data in tables and graphs	To build pie charts, column charts, histograms and frequency polygons
	To collect, organise and describe data in tables and graphs	Absolute and relative frequency samples, population
Central tendency measures	To understand mean as a tendency indicator	To calculate and understand mean, mode, median
Sample space, probability	Synthesise, communicate and draw conclusions	To present the data globally, highlighting relevant aspects, to allow inferences
	Sample space and estimation by ratio	Multiplicative principle Design of experiments and simulations to estimate probabilities and check provided probabilities

Franklin and Kader (2010) propose that the knowledge required by math teachers who are teaching statistics must be based upon three principles: knowledge of what we would expect educated members of society to know; that the knowledge used for teaching statistics is not the same as the statistical knowledge needed for other statistics-based professions; and, the statistical knowledge needed for teaching must be usable for such challenges as interpreting a student's error, using multiple forms to represent a statistical idea and developing alternative explanations.

Thus, the Brazilian teachers who teach statistics need to understand the basic concepts of statistics. They should gain “both technical and conceptual knowledge” of the statistics and probability content that appears in the curriculum for their students, which is described in Brasil (1998) and is summarised in Table 33.1.

In this context, we have, in our study, similar goals to Peck and Gould (2005, p. 1): the development of an experience for pre-service teachers that would provide them with the necessary background to teach introductory statistics; the provision of help for teachers to incorporate real data, active learning and technology in teaching introductory statistics; to provide access to a variety of resources for teaching statistics; to create a community of learners who can advise and support each other in matters of classroom practices, pedagogy and understanding statistical concepts.

To achieve these goals, we trained pre-service teachers in applying two Didactical Sequences (DS): “Profile of the Class” and “Water Planet” (Cazorla & Santana, 2010).

The concepts involved in these DS are suggested in “Treatment of Information” (data handling) Section (Brasil, 1998). These DS were designed for statistics teachers and are available in the Virtual Environment to Support Statistics Literacy Basic Education: AVALE-EB (Fig. 33.1).

In these DS students follow the steps of a cycle of scientific investigation: background of the problem situation, formulation of hypotheses, defining the variables, collecting and recording the data, calculation of statistical measures and constructing tables and graphs, data interpretation and communicating the results.



Fig. 33.1 AVALE-EB's homepage (<http://www.iat.educacao.ba.gov.br/avaleeb>)

These DS use interdisciplinary content, contextualised within the school environment, exploring the cognitive aspects of learning and those related to the development of critical awareness of the use of natural resources and a respect for diversity. In DS students actively participate in all activities, so DS can contribute to statistical literacy by using statistical and probabilistic concepts that enable the development of the critical reading ability of basic education students and can also contribute to the scientific education of these students.

33.3 Methodology

Initially, the four pre-service teachers were invited to participate in a training activity on two Didactical Sequences (DS) that would work with students in middle schools.

Following the training activity the four pre-service teachers made a plan to implement each DS at school for 2 h per week for a month, using 3 weeks to work in a “pencil and paper environment” and 1 week to work in a “computing environment” followed by the application of a small statistical test. These activities were developed with students from five classes of sixth and seventh grade, and three classes of eighth and ninth grade students, totalling 16 classes in middle school.

33.4 Key Results

Throughout their training we found that despite the pre-service teachers having already attended an introductory course in statistics, they didn't know how to identify the nature of variables, which is an important competency to work within the DS "Profile of the Class", in which students are encouraged to formulate research questions and helped to identify the treatment of nominal or ordinal qualitative variables and discrete or continuous quantitative variables and shown how to represent their data. They also revealed a feeling of insecurity, due not so much to lack of preparation in statistics, but more especially to a lack of preparation for the teaching of statistics. As a result the pre-service teachers studied each step of the sequences and did additional reading to remedy their concerns and develop complementary activities before implementing DS with the students, especially in the interpretation of graphs for middle school students, as this is the most commonly required activity in Brazilian educational assessments.

In the first application of the DS "Profile Class" with the classes of seventh grade, students were very excited and formulated several questions about aspects of everyday life, regarding for instance, favourite soccer team, number of pets, favourite food, taste in vegetables and salads, favourite sports and favourite teacher. The DS gave students the opportunity to learn statistics while working with their own data and the contents explored could include: charts: bar, pie and dotplot graphs, measures of central tendency (mean, median and mode) and measures of dispersion (total amplitude).

In the "pencil and paper environment", the activities were as planned: students were asked to construct statistical tables or graphs, they could see how to build up and to understand information from different kinds of sources. One of the pre-service teachers felt that there was a need to strengthen the interpretation of several kinds of chart in his two classes.

Before working with measures of central tendency with the students' data, the pre-service teachers decided together that it would be interesting if such measures were intuitively perceived and they decided to use a game.

The game chosen was based on Super Trunfo[®] (Super Trumps) because it requires throughout that the player realises that there are better cards than others in certain respects, and values that tie more easily than others. The original game has 32 cards and each card contains 6–8 pieces of information.

The pre-service teachers decided to create a version of Super Trunfo[®] themed movies with fewer cards (20), separated into four groups of five. The game produced was named "Super Movie". It was important to reduce the number of cards and the amount of information from that in the original game so that study of the distribution of data was possible.

The pre-service teachers also studied the dynamics of watching games with students: how the cards were distributed according to three variables (release year,



Fig. 33.2 Examples of cards designed by pre-service teachers for the game “Super Movie”

duration and revenue) and the best strategy for certain groups of cards (Fig. 33.2). Throughout the game the students realised for example, that revenues worth more than \$1 billion had great chances of winning and that movies of more than 120 min duration were too long. The median, which is traditionally studied as an end in itself, had a practical utility in the game, because if in one of the variables, a piece of information was greater than the median, the possibility of gaining the card was over 50 %. So the students had to compare variables, and calculate the mean, median and mode to beat their opponent. This game stressed statistical concepts.

The pre-service teachers tried to use the “computing environment” with students, but there were problems, for example a slow internet connection when accessing the AVALE-EB made it impossible to work with the DS.

In the statistical test these students performed well. The pre-service teachers believe that their performance indicated that the questions were appropriate to the topics covered in this DS.

In the application of the DS “Water Planet”, with students from the eighth and ninth grades, the concepts were studied only in the “pencil and paper environment”. The contents that can be explored are: variables ordered by time; charts: bar, line and dot plot; measures of central tendency (mean, median and mode) and measures of dispersion (total amplitude, deviation, mean deviation, variance, standard deviation and coefficient of variation). It is also possible to relate the concept of per capita consumption of water with the arithmetic mean.

The pre-service teachers expected that using data from the student’s own water bills (Fig. 33.3) would motivate them to work with this DS. However this didn’t happen, as most of the students didn’t provide their own water bills, requiring pre-service teachers to work with data from the AVALE-EB database (Fig. 33.4).

		Data de Vencimento 10/02/2010		FATURA Nº. 2a. VIA									
Nome Severino Silva		Código do Consumidor 760.724-4		Data de Emissão 05/02/2010									
Endereço do Usuário Rua Castro Alves		Nº Porta 934		Bairro CENTRO									
Endereço de Entrega		Nº Porta		Bairro									
Categoria-Economias RNP-001 -		Localização 01.001.0076.02.0934.00		Mês de Referência 01/2010									
Data de Leitura 03/02/2010	Leitura atual 1.694	Leitura anterior 1.686	Consumo m³ 8	Número do Hidrômetro A95T027799									
Consumo dos últimos meses	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	Média dos últimos meses
	7	11	8	9	9	8	6	7	9	10	15	8	8,9

Fig. 33.3 Example of a water bill in Brazil

Nome	Jan	Feb	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
aida	115	103	142	177	131	115	82	114	126	119	126	126
amanda	196	178	187	185	156	159	153	168	163	161	186	186
brandao	220	180	230	260	24	235	268	245	250	238	256	220
cido	115	103	142	177	131	115	82	114	126	119	126	126
cris	288	314	311	414	357	370	399	350	350	367	326	338

Fig. 33.4 Class database in AVALE-EB

This activity demonstrated the difficulty students had in working with scales and the “simple rule of three”. More time was thus spent making and interpreting graphs. The third week began with definitions of some measures of dispersion, amplitude, deviations and mean deviation. These measures were presented using graphs previously used by the students to help them understand their meaning. For example, students were asked to chart the average monthly water consumption of a family using a bar graph showing the consumption each month, and this showed that the deviations were nothing more than the distance from the top of the base to the mean line.

Students were given a table to fill with the data provided in the water bills and with which to calculate several measures of dispersion (Fig. 33.5). With this table the students made calculations of the dispersion measures discussed and were able to make comparisons with their peers.

In the second school the pre-service teachers decided not to use AVALE-EB for the calculations, and instead made a simple table in Excel format, similar to that given to the students to fill in at the third stage. The table was completed automatically when students entered the water consumption for each month. The graphics were also constructed automatically when the table was completed.

The statistical test was very extensive and demonstrated that the concepts of median and average deviation weren’t learned effectively in both schools.

Month(i)	Consumption (x_i)	Mean* (\bar{x})	Deviation ($x_i - \bar{x}$)	Square Deviation ($(x_i - \bar{x})^2$)
January				
February				
March				
April				
May				
June				
July				
August				
September				
October				
November				
December				
Sum		-----		
Mean		-----		
*To repeat the mean in the all lines			↑ Mean Deviation	↑ Variance

Fig. 33.5 Example of worksheet for the ST planet water

33.5 Final Considerations

The pre-service teachers were limited in several ways: they were most focused on pedagogy over content. It considered that training activity as short workshop format is not conducive to the development of full content understanding. This is in according to Gattuso and Pannone (2002).

The difficulties experienced by the pre-service teachers can be explained, in part because the subject is not covered in their initial course for teachers of mathematics, and they are therefore untrained to present the contents of the Data Handling project and the didactic aspects that should be taught in basic education.

Their initial course for teachers of mathematics have only a list of concepts and procedures, but there is no evidence of statistics as a tool of quantitative scientific research, which allows the formulation of hypotheses, planning the collection, processing and analysis of data, nor as a language that permeates information conveyed by the media.

In general, the experiment in school was evaluated positively, due to the active participation of the students—capable students took responsibility for helping the weak students, lazy students were stimulated to work—and because most students were able to make sense of the tables and graphs, and learned to compute the mean, the median, mode for much grouped data with the help of Excel or AVALE-EB.

After qualitative correction of the statistical tests, the pre-service teachers discussed the results amongst themselves and with their advising professors and

supervising teachers about, considering perceptions of failure and changes to the next application of these DS. For the in-service teachers the PIBID has been an opportunity for continued education. One of the supervisors, who had been in a state of great emotional distress about the profession, and had demonstrated great apathy in merely repeating former semesters, changed his attitude with the new activities proposed by the pre-service teachers and the reasoned response of the students, going to give suggestions and participate in more activities in general school.

It is therefore believed that the PIBID has provided important moments of training to pre-service teachers: in teaching statistical concepts that will be needed for teaching in the future, and providing opportunities to prepare and execute activities in the classroom with the guidance and supervision of professors and teachers, which gives greater security for pre-service teachers who are learning to teach.

Acknowledgments CAPES (PIBID ICMC-USP) and CNPq (472295/2010-1) for financial support.

References

- Batanero, C., Godino, J. D., & Roa, R. (2004). Training teachers to teach probability. *Journal of Statistics Education*, 12. <http://www.amstat.org/publications/jse/>
- Brasil, Ministério da Educação, Secretaria de Educação Fundamental. (1997). *Parâmetros Curriculares Nacionais: matemática*. Brasília: Ministério da Educação/Secretaria de Educação Fundamental.
- Brasil, Ministério da Educação, Secretaria de Educação Fundamental. (1998). *Parâmetros Curriculares Nacionais: matemática*. Brasília: Ministério da Educação/Secretaria de Educação Fundamental.
- Brasil, Ministério da Educação. Secretaria de Educação Média e Tecnológica. (2002). *PCN Ensino Médio: Orientações Educacionais complementares aos Parâmetros Curriculares Nacionais— Ciências da Natureza, Matemática e suas Tecnologias*. Brasília: Ministério da Educação.
- Brasil, Ministério da Educação. Secretaria de Educação Média e Tecnológica. (2006). *Orientações curriculares nacionais para o Ensino Médio. Ciências da Natureza, Matemática e suas Tecnologias*. Brasília: Ministério da Educação.
- Canada, D. (2006). Elementary pre-service teachers' conceptions of variation in a probability context. *Statistics Education Research Journal*, 5(1), 36–63.
- Cazorla, I. M. (2006). Teaching statistics in Brazil. In: A. Rossman & B. Chance (Eds.), *Proceedings of Seventh International Conference on Teaching Statistics*. Intern. Assistant for Statist Education, Salvador (Brazil). CD-ROM.
- Cazorla, I. M., & Santana, E. (Orgs.). (2010). *Do Tratamento da Informação ao Letramento Estatístico*. Itabuna: Via Litterarum.
- Contreras, J. M., Batanero, C., Díaz, C., & Fernandes, J. A. (2011). Prospective teachers' common and specialized knowledge in a probability task. In T. Rowland et al. (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education*. Polónia: University of Rzeszów.
- Franklin, C., & Kader, G. (2010). Models of teacher preparation designed around the GAISE framework. In: C. Reading (Ed.), *Proceedings of Eighth International Conference on Teaching Statistics*. Intern. Assistant for Statist Education, Ljubljana (Slovenia). CD-ROM.
- Gattuso, L., & Pannone, M. (2002). Teacher's training in a statistic teaching experimentation. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa: International Statistical Institute and International Association for Statistical Education. <http://www.stat.auckland.ac.nz/~iase/publications>.

- Groth, R. E., & Bergner, J. A. (2004). Pre-service elementary school teachers metaphors for the concept of statistical sample. *Statistics Education Research Journal*, 4(2), 27–42.
- Leavy, A. S. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning. *Statistics Education Research Journal*, 9(1), 46–67.
- Nichololson, J., Road, C., & Darnton, C. (2003). Mathematics teachers teaching statistics: What are the challenges for the classroom teacher? In *Proceedings of the 54th Session International Statistical Institute, Berlin*. Alemanha: International Statistical Institute.
- Peck, R., & Gould, R. (2005). Preparing secondary teachers to teach statistics: A distance education model. *Invited paper at the International Statistical Institute 55th Session*, Sydney. <http://www.stat.auckland.ac.nz/~iase/publications>.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: Macmillan.
- Viali, L. (2008). O ensino de Estatística e Probabilidade nos cursos de licenciatura em Matemática. *Anais do 18º Simpósio Nacional de Probabilidade e Estatística. Estância de São Pedro, SP: Associação Brasileira de Estatística*.

Chapter 34

Statistics in Primary Education in Greece: How Ready Are Primary Teachers?

Eugenia Koleza and Aristoula Kontogianni

34.1 Introduction

Although there is a prominent belief that statistics is not actually mathematics (e.g., Cobb & Moore, 1997; delMas, 2004; Pereira-Mendoza, 2002), in the school reality is part of the mathematics syllabus. In Greece, like other countries worldwide (e.g., NCTM, 2000 for USA; [ACARA], 2010 for Australia) statistics is taught within the course of mathematics.

Compulsory education in Greece consists of primary and secondary education and its duration is of 9 years. Primary education is about primary school which consists of six grades (ages 6–12). Statistical notions that were taught in the last three grades according to the national standards (CTCF, 2003) included:

- Collection, organizing, and interpretation of data.
- Construction and interpretation of graphs (bar graphs, pictographs, and line graphs).
- Computation and interpretation of the mean.

The new curriculum that was released on May 2011 places more emphasis on statistics: statistics will be taught in every grade and will include, in addition to the previous curriculum, the teaching of stem-and-leaf plots, line plots, double bar graphs, mode, median, and range. Given that a recognized key player in the

E. Koleza (✉) • A. Kontogianni
University of Patras, Patras, Greece
e-mail: ekoleza@upatras.gr; akontog@upatras.gr

curriculum reform process is the teacher (Shaughnessy, 1992), it turns out that successful implementation of new statistical contents will be problematic because of poor knowledge and inexperience of primary teachers.

34.2 Research About Teachers' Statistical Knowledge

While there are many studies about the statistical competence of students (e.g., Garfield, 2003; Watson & Callingham, 2003), those concerning statistical competence of teachers are relatively few and concern mainly: comprehension of measures of center and graphs as well as teachers' beliefs and attitudes towards statistics. More precisely, studies about teachers' comprehension of measures of center (e.g., Batanero, Godino, & Navas, 1997; Cai & Gorowara, 2002; Groth & Bergner, 2006; Jacobbe, 2008) revealed a lack of understanding of the algorithm for calculating the average, a difficulty in discriminating the mean with the other measures of center and little or no understanding of the effect of outliers on the mean (Jacobbe & Carvalho, 2011).

Research involving teachers' comprehension of graphs (e.g., Bruno & Espinel, 2009; Espinel, Bruno, & Plasencia, 2008; Jacobbe & Horton, 2010; Monteiro & Ainley, 2003) revealed that teachers had difficulties in the interpretation of statistical graphs, they weren't able to make generalizations about the data (Gonzalez et al., 2011) and were unsuccessful with questions that assessed higher levels of graphical comprehension (Jacobbe & Horton, 2010).

Studies about teachers' beliefs and attitudes towards statistics (e.g., Begg & Edwards, 1999; Chick & Pierce, 2008; Estrada et al., 2008) showed that they reacted in a negative way towards the subject in question, although they stated that statistics enables people to comprehend better their world.

In Greece, research (Chadjipadelis, 1999; Page, 1999) is restricted in results' analysis of professional development programs, without a prior examination of the statistical knowledge level of participants. The present study, given the appearance of new mathematics curricula, attempts to investigate and describe the level of statistical competence of primary teachers in Greece.

34.3 Theoretical Framework and Research Question

Several studies have shown that teacher knowledge is connected to what and how students learn and depends on the context in which it is used (Ball & Bass, 2000; Cobb, 2000). According to Shulman (1986), there are three layers of knowledge that are necessary for effective teaching: (a) subject matter knowledge, (b) pedagogical matter knowledge, and (c) curricular knowledge. Our conviction is that in order to teach according to modern standards, teachers need to understand subject matter deeply and flexibly enough, so they will be able to help students create useful cognitive maps, relate one idea to another, and address misconceptions. Teachers need to see how mathematical ideas connect across different fields and to everyday

life. This kind of understanding provides a foundation for pedagogical content knowledge that enables teachers to make ideas accessible to others (Shulman, 1987). This is why in our research we focused on teachers' subject matter knowledge. More specifically, our research question was as follows:

How ready are primary school teachers in Greece for the teaching of statistics with respect to the standards set forth in the new curriculum of mathematics?

The research question was restricted to the knowledge of primary teachers in regard to:

- The computation of measures of center and measures of spread. The reading, interpreting, and inferring data from graphical displays.
- The above competencies were chosen because of their importance as they are part of the mathematics curriculum of primary school and they are omnipresent in research literature.

34.4 Method

In order to answer the research question, a combination of quantitative and qualitative techniques was used. The primary researcher spent extensive time in the field of reading and analyzing the official documents and similar studies, constructing and administrating an assessment instrument, and conducting interviews with the participants.

34.4.1 Setting and Participants

The participants were selected using purposeful sampling (Patton, 1990). The ten participants (four female and six male) were selected because they were rated highly as they were all principals and vice-principals of primary schools in an urban area located in the southwest region of Greece. In particular, these teachers were selected as they had many years of teaching experience (from 10 to 32 years) in all grades of primary education and all had participated in programs of professional development. Three of them had a master degree in teaching science, one a doctorate degree in philosophy and two a second degree in early childhood education.

The research was conducted during the first days of September 2011 as September is the first month of school year and the new standards for Mathematics education were released on May of 2011.

34.4.2 Two Phases of Data Collection

The data for the present study were collected during two phases. In the first phase, an assessment instrument was administrated to the participants. The second phase of data collection consisted of face-to-face semi-structured interviews, duration of 20 min, with each one of the participants. The interview's purpose was to clarify

individual teacher responses of the written instrument and for this reason the questions were guided by the nature of the written responses.

The assessment instrument was developed to assess only statistical content knowledge with respect to the standards as proposed by the official documents. Reliability of the instrument was tested using a Cronbach's alpha simulation approach (Leontitsis & Pagge, 2006) and the results showed $\alpha=0.83$.

The assessment instrument's items in order to measure the statistical content knowledge included the following:

- Reading, interpreting, and inferring data from graphical displays such as line plots, stem-and-leaf plots, pictographs, and double bar graphs.
- Computation of measures of center (mean, median, and mode) and measures of spread (range) and inferring conclusions from these.

The assessment instrument measured different levels of cognitive outcomes for each aspect and consisted of seven open-ended items with multiple parts intended to be answered in 45 min. Statistical literacy's domain was measured through the recognition, identification, computation, and understanding of measures of center and spread using the framework developed by Garfield (2002) and delMas (2002). As for the graphical displays, the three components of graph comprehension were considered: reading the data, reading between the data, and reading beyond the data (Curcio, 1987; Friel, Curcio, & Bright, 2001).

The analysis was conducted by coding each item's responses utilizing a rubric for levels of correctness (0–4) adapted from Garfield (1993) and Thompson and Senk (1998) and used by Sorto (2004). The total number of questions was 17, making a total of 68 possible points.

34.4.3 Limitations

The present study is limited by the sample size because examining only ten teachers is inappropriate to infer for all primary teachers. Nonetheless, because of the fact that for the data collection it was used an assessment instrument followed by interviews and the sample consisted of ten highly ranked and well-educated teachers allows us to make an early estimation about the level of statistical content knowledge of the primary in-service teachers in Greece.

34.5 Discussion and Results

34.5.1 Overall Performance

In general, the participants' performance was not sufficient. Specifically, as it concerns the first part of our research question—about the measures of central tendency—the majority of them (9–10) were unable to compute the median for a set of

ten numbers, as they were unfamiliar with this concept. They interpreted “median” as “the number in the middle” when the data are placed in order, but they didn’t know how to compute it. All of them were able to compute the mean, but four of them answered that it can’t be a decimal number. Despite the fact that they were able to compute the mean in various contexts (data sets or graphs), most of them had difficulties to understand that it is an indication point for a data set rather than a number that can give as always accurate information. However, it is amazing that two of them noticed that the mean can’t be representative enough of a set of data when there are outliers, without having prior knowledge of it. As it concerns the notion of range, all of them were able to understand what it is about, but described it as a space rather than a number.

In respect to the second part of our research question, the results showed that they were able to understand and interpret double bar graphs and pictographs, but they had difficulty in understanding the structure of the stem-and-leaf plot and of the line plot. These particular graphs are new to mathematics curriculum and most of the teachers believed that are inappropriate for the primary school. In addition, most of them connected the choice of a graph with the conclusion that occurs and not with the data that are represented by the particular graph.

34.5.2 Discussion of Specific Items

The majority of the participants showed difficulty in answering the questions that involved a stem-and-leaf plot and a line plot, although an explanation was provided. These items come from the research of Sorto (2004) (Figs. 34.1 and 34.2).

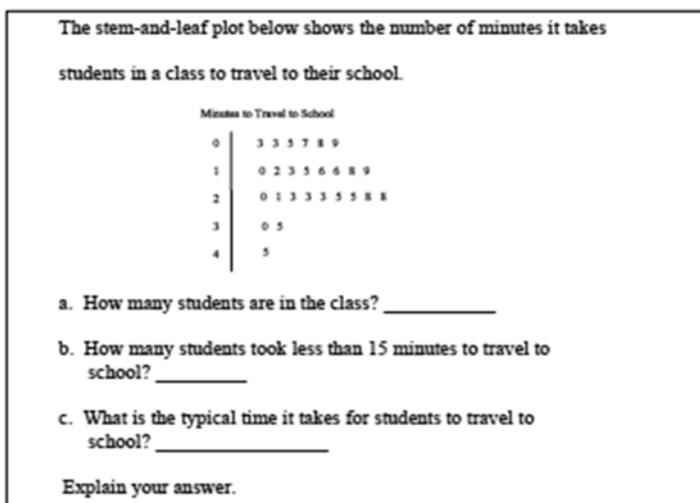


Fig. 34.1 Stem-and-leaf plot question (Sorto, 2004)

Specifically for the stem-and-leaf plot teachers expressed their disapproval as they believed that it was very difficult for them and for their students. They supported their belief by the fact that they weren't able to make inferences about it, as they showed difficulty in combining the numbers at the two sides of the stem.

Interviewer: What do you think about the stem-and-leaf plot?

Sophia: It is very difficult... I can't understand it...I can't believe that my students will be able to understand it...

Interviewer: Why?

Sophia: Because of the way that the data are represented...The fact that there is a digit that you must combine with the rest numbers is very confusing...I believe that it will be very difficult for the students...some of them aren't able to discriminate the value of the digits in a number...

Interviewer: What do you think about the stem-and-leaf plot?

Maria: It is very difficult for me... It is the first time that I see such a graph... I couldn't "read" it because I was not able to understand the connection between the numbers at the two sides of the line...

The next item was about a line plot and the notion of the mean. Except for the difficulty they showed in understanding of the certain graph, they expressed also misunderstandings about the mean (Fig. 34.2).

Interviewer: Is it possible the mean to be a decimal number?

Bill: No it can't be, because we use the mean in order to express a value approximately.... So it isn't necessary for us to be punctual. It isn't necessary to use decimal numbers.

Interviewer: Do you think that the mean can be a decimal number?

Ann : I don't know...Is it possible to have a family with 3.5 members?...It doesn't exist such a thing...No, I don't think so...but I can't tell for sure...

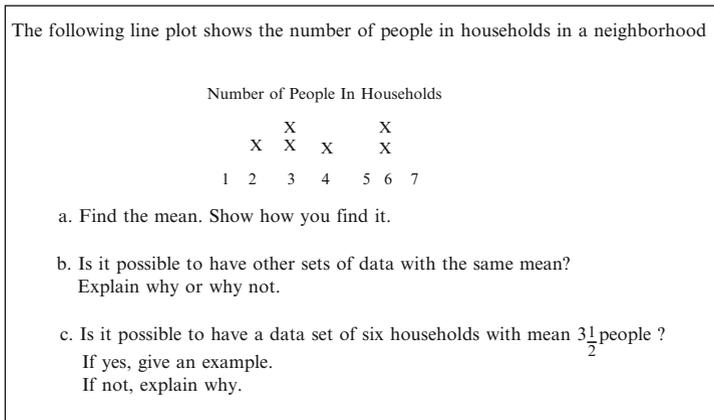


Fig. 34.2 Line-plot question (Sorto, 2004)

Ten students take every day as allowance the above amount of money
 4,5,3,2,2,7,8,4,5,2
 For these data compute the median, the mode and the range.

Fig. 34.3 Measures of center and spread (Sorto, 2004)

Among the items concerning the measures of center and spread, the following one that was constructed by the researchers proved to be difficult for the teachers as they lacked the necessary content knowledge: Not only most of them didn't know how to compute the median, furthermore, the analysis of the transcript showed that they confused it with the mean and the mode (Fig. 34.3).

Interviewer: What is the median?

Sophia: It is the value that occurs more often...the repeated value...

Interviewer: Is it the same thing the mean and the median?

Andrew: No, it isn't ...The median is the tension I think...the tension that exists in the data. The mean shows us an element...that...I can't describe it...it shows us the middle of our data while the median shows the tension....

34.6 Conclusion

The main conclusion of this study is that primary teachers involved had a low-level knowledge of basic statistical notions. Some of these notions like the median, the mode, the range, the stem-and-leaf plot, and the line plot were completely unknown to them.

This result coincides with the results of previous studies (Jacobbe & Horton, 2010) as it is impossible for people to have mastery of ideas that have not been taught. However, this study also revealed that the participants were unable for deeper thinking even in notions that they were familiar with, like the mean or the choice of the appropriate graph. This fact leads us to conclude that only content knowledge is not sufficient and the implementation of professional development programs for the in-service teachers in the field of statistics is necessary. If the main goal of a mathematics curriculum is not only superficial knowledge for students, but a deep understanding of statistical notions, it is important for the teachers first to understand these notions at least one level beyond the level they will teach. If teachers strive to make their students lifelong learners, they too must continue to be lifelong learners, for this is where real understanding takes place (Arnold, 2008). With adequate training, teachers will be more confident and they will be able to encourage students to speculate and explore phenomena, to create their own data representations and make their own conjectures instead of limiting them to the practice of procedural skills and execution of calculations (Gattuso & Ottaviani, 2011).

The partial results of this study—though limited—coincide with studies around the world (e.g., Groth & Bergner, 2006; Jacobbe & Horton, 2010; Sorto, 2004) about the teachers' low level in understanding statistics. As it concerns the measures

of center, it is profound that the understanding of the content needed to teach these notions at the elementary level is a nontrivial matter (Groth & Bergner, 2006). While, as far as graphs are concerned, it was apparent that the in-service teachers had a low-level comprehension of these (Jacobbe & Horton, 2010). The results of this study reveal that in Greece, like other countries around the world, the teaching of statistics needs a different approach in contrast to other subjects in mathematical syllabus.

Acknowledgments This research has been co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF)—Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

References

- ACARA. (2010). Australian Curriculum Consultation Portal.
- Arnold, A. (2008). Developing new statistical content knowledge with secondary school mathematics teachers. In C. Batanero, G. Burril, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 Conference and IASE 2008 Round Table Conference*. Monterrey: ICMI and IASE.
- Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 83–104). Westport, CT: Alex.
- Batanero, C. Godino, J., & Navas, F. (1997). Concepciones de maestros de primaria en formación sobre promedios [Primary school teachers' conceptions on averages]. In H. Salmerón (Ed.), *Actas de las VII Jornadas LOGSE: Evaluación Educativa* (pp. 310–340). Granada, Spain: University of Granada.
- Begg, A., & Edwards, R. (1999, December). Teachers' ideas about teaching statistics. *Paper presented at the Joint Conference of the AARE and NZARE*, Melbourne.
- Bruno, A., & Espinel, M. C. (2009). Construction and evaluation of histograms in teacher training. *International Journal of Mathematical Education in Science and Technology*, 40(4), 473–493.
- Cai, J., & Gorowara, C. C. (2002). Teachers' conceptions and constructions of pedagogical representations in teaching arithmetic average. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society, Cape Town, South Africa*. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute. [http://www.stat.auckland.ac.nz/~iase/publications/1/2c1_cai.pdf].
- Chadjipadelis, T. (1999). Teaching teachers to teach statistics. *Contributed Paper at the International Statistical Institute*, 52nd Session, Helsinki.
- Chick, H., & Pierce, R. (2008). Teachers' beliefs about statistics education. In C. Batanero, G. Burril, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 Conference and IASE 2008 Round Table Conference*. Monterrey: ICMI and IASE. [www.stat.auckland.ac.nz/~iase/publications/rt08/T2P13_Chick.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: Alex.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly*, 104(9), 801–823.

- Cross Thematic Curriculum Framework and Curricula (CTCF) (2003). Government Paper 303/13-3-2003, Volume II, pp. 3733-4068 & Government Paper 304/13-3-2003, Volume II, pp. 4069-4400 (in Greek).
- Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18(5), 382-393.
- delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79-96). Dordrecht, The Netherlands: Kluwer.
- delMas, R. C. (2002). Statistical literacy, reasoning and learning: A commentary. *Journal of Statistics Education*, [Online] 10(3).
- Espinell, M. C., Bruno, A., & Plasencia, I. (2008). Statistical graphs in the training of teachers. In C. Batanero, G. Burril, C. Reading, & A. Rossman (Eds.), Joint ICMI/IASE study statistics in school mathematics. Challenges for teaching and teacher education. *Proceedings of the ICMI Study 18 Conference and IASE 2008 Round Table Conference*. Monterrey: ICMI and IASE. [www.stat.auckland.ac.nz/~iase/publications/rt08/T2P13_Espinell.pdf].
- Estrada, A., et al. (2008). Teachers' attitudes towards statistics. In C. Batanero, G. Burril, C. Reading, & A. Rossman (Eds.), Joint ICMI/IASE study statistics in school mathematics. Challenges for teaching and teacher education. *Proceedings of the ICMI Study 18 Conference and IASE 2008 Round Table Conference*. Monterrey: ICMI and IASE. [www.stat.auckland.ac.nz/~iase/publications/rt08/T2P13_Estrada.pdf].
- Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124-158.
- Garfield, J. B. (1993). An authentic assessment of students' statistical knowledge. In N. L. Webb & A. F. Coxford (Eds.), *Assessment in the Mathematics Classroom* (pp. 187-196). Reston, VA: 1993 Yearbook of the National Council of Teacher of Mathematics (NCTM).
- Garfield, J. B. (2002). The challenge of developing statistical reasoning. [Online] *Journal of Statistics Education*, 10(3).
- Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38.
- Gattuso, L., & Ottaviani, M. G. (2011). Complementing mathematical thinking and statistical thinking in school mathematics. In C. Batanero, G. Burril, & C. Reading (Eds.), Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICME/IASE study. New York: Springer.
- Gonzalez, M. T., et al. (2011). Teachers' graphical competence. In C. Batanero, G. Burril, & C. Reading (Eds.), Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICME/IASE study. New York: Springer.
- 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.
- Jacobbe, T. (2008). Elementary school teachers' understanding of the mean and median. In C. Batanero, G. Burril, C. Reading, & A. Rossman (Eds.), Joint ICMI/IASE study statistics in school mathematics. Challenges for teaching and teacher education. *Proceedings of the ICMI Study 18 Conference and IASE 2008 Round Table Conference*. Monterrey: ICMI and IASE. [www.stat.auckland.ac.nz/~iase/publications/rt08/T2P13_Jacobbe.pdf].
- Jacobbe, T., & Carvalho, C. (2011). Teachers' understanding of averages. In C. Batanero, G. Burril, & C. Reading (Eds.), Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICME/IASE study. New York: Springer.
- Jacobbe, T., & Horton, R. (2010). Elementary school teachers' comprehension of data displays. *Statistics Education Research Journal*, 9(1), 27-45.
- Leontitsis, A., & Pague, J. (2006). A simulation approach on Cronbach's alpha statistical significance. *Mathematics and Computers in Simulation*, 73, 336-340.
- Monteiro, C. & Ainley, J. (2003). Developing Critical Sense in Graphing. In *Proceedings of III Conference of the European Society for Research in Mathematics Education*. http://www.dm.unipi.it/~didattica/CERME3/proceedings/Groups/TG5/TG5_list3.html.

- National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: National Council of Teachers of Mathematics.
- Page, J. (1999). How to teach statistics to a life-long learning group of school teachers. *Contributed Paper at the International Statistical Institute 52nd Session*, Helsinki. [www.stat.auckland.ac.nz/~iase/publications/5/pang0346.pdf].
- Patton, M. (1990). *Qualitative evaluation and research methods*. Beverly Hills, CA: Sage.
- Pereira-Mendoza, L. (2002). Would you allow your accountant to perform surgery? Implications for education of primary teachers. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS 6)—Cape Town* (pp. CD Rom). Voorburg, The Netherlands: International Association for Statistical Education.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 623–660) Reston, VA: National Council of Teachers of Mathematics.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- Sorto, M. A. (2004). Prospective middle school teachers' knowledge about data analysis and its application to teaching. Unpublished doctoral thesis, Michigan State University.
- Thompson, D., & Senk, S. (1998). Using rubrics in high school. *The Mathematics Teacher*, 91(9), 786–793.
- Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(3), 3–46.

Chapter 35

Developing Statistical Literacy (DSL): Student Learning and Teacher Education

Hélia Oliveira, Ana Henriques, and João Pedro da Ponte

Recent mathematics curriculum orientations stress the importance of developing students' statistical literacy at different school levels. Many students can read and understand tables, charts, and graphs and perform the procedures to find statistical measures, but they miss the conceptual abilities to interpret and draw conclusions from graphs, or to make decisions on which calculation is appropriate to study a particular situation (Shaughnessy, 2007). In Portugal, the mathematics syllabus for basic education which started in 2010 presented more demanding learning objectives for statistics. This was a challenging situation for teachers requiring them to develop new perspectives and professional knowledge, since many of them did not have a suitable preparation in statistics. In this context, we planned a project aiming to construct knowledge about statistical literacy development, with two main foci: (1) the characterization of students' statistical literacy from preschool to secondary levels, and the possibilities and constraints for its development, and (2) the development of teachers' statistical and didactical knowledge for teaching in in-service and preservice teacher education settings. A review of literature supported both the planning and development of teaching experiments in classrooms at different school levels, based on sequences of tasks and the use of technological tools and the planning of courses in preservice education of educators and elementary school teachers and in-service teacher education, as well as creating contexts of collaborative work with teachers. The project used a design research methodology in teaching and teacher education experiments. Data was collected by a variety of methods including interviews, classroom observations, students' written work, tests, and questionnaires. Data analysis included mainly qualitative approaches.

H. Oliveira (✉) • A. Henriques • J.P. da Ponte
Instituto de Educação, Universidade de Lisboa, Lisboa, Portugal
e-mail: hmoliveira@ie.ul.pt; achenriques@ie.ul.pt; jpponte@ie.ul.pt

This study was supported by national funds through FCT – Fundação para a Ciência e Tecnologia, Project Developing statistical literacy: student learning and teacher training (contract PTDC/CPE-CED/117933/2010).

Reference

- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Greenwich, CT: Information Age and NCTM.

Chapter 36

Teacher Capacity as a Key Element of National Curriculum Reform in Statistical Thinking: A Comparative Study Between Australia and China

Qinqiong Zhang and Max Stephens

36.1 Introduction

In the official curriculum documents of many countries, statistics and statistical reasoning have become part of the mainstream in school curriculum. In The Australian Curriculum: Mathematics (ACARA, 2010), “Statistics and Probability” is one of three key content areas. In its overview statement to this strand, ACARA (2010) states that: “Statistics and probability initially develop in parallel, and the curriculum then progressively builds the links between them. Students recognise and analyse data and draw inferences...They develop ... to critically evaluate chance and data concepts ...and develop intuitions about data” (p. 2). A corresponding strand, Chance and Data has been present, for at least 5 years, in related State curriculum documents. e.g. VELS (VCAA, 2008) and the Mathematics Developmental Continuum (DEECD, 2006). China’s newly revised Mathematics Curriculum Standard for Compulsory Education (Ministry of Education of PRC, 2011) also presents a single strand entitled Statistics and Probability. In the overall objective for this content strand, it is stated that “to experience the process of collecting and dealing with data in practical problems, as well as using data to analyse questions and obtaining information” in “knowledge and skills” (p. 8); and it refers to “to experience the significance of statistical methods, to develop ideas of statistical analysis and to experience random phenomena” in “mathematical thinking” (p. 9).

Q. Zhang (✉)
Wenzhou University, Wenzhou, China
e-mail: qqzhang922@126.com

M. Stephens
The University of Melbourne, Parkville, VIC, Australia
e-mail: m.stephens@unimelb.edu.au

These intentions are endorsed by Garfield and Ben-Zvi (2008) who point out that in contrast to traditional approaches to teaching which focus on computations of theoretical probability, new emphases are squarely focused on understanding data and development of statistical thinking and literacy (p. 7). They argue that “the goals for students at the elementary and secondary level tend to focus more on conceptual understanding and attainment of statistical literacy and thinking and less on learning a separate set of tools and procedures” (p. 14). These goals are reflected in the National Curriculum in Australia and China, where students are expected to learn and understand that: (1) explanations supported by data are more powerful than personal opinions or anecdotes; (2) variability is natural and is also predictable and quantifiable; (3) association is not the same as causation; and (4) random sampling allows results of surveys and experiments to be extended to the population from which the sample was taken. (cf. Garfield & Ben-Zvi, 2008, p. 15).

However, the implementation of curriculum change is never simply from the top down. Teachers’ interpretations and responses at the level of practice are never simple reflections of what is contained in official curriculum documents. While curriculum documents set out broad directions for change, any successful implementation of these “big ideas” depends on teachers’ capacity to apply subtle interpretations and careful local adaptations (Datnow & Castellano, 2001). We argue that teacher capacity is a key dimension in realising that goal.

36.2 Teacher Capacity and Mathematical Knowledge for Teaching

While the term “teacher capacity” is not widely used in mathematics education research, it has clear connections with the research of “Pedagogical Content Knowledge” by Shulman (1986, 1987) and “Mathematical Knowledge for Teaching” by Ball, Thames and Phelps (2008).

36.2.1 *Shulman’s Model*

Shulman (1987) identified *pedagogical content knowledge* as the category most likely to distinguish the understanding of the content specialist from that of the expert teacher. The importance given to PCK suggests that what is needed in mathematics teaching is not just knowledge of the subject, or general knowledge of pedagogy, but rather a combination of both. However, after 25 years of exposure to Shulman’s thinking, Petrou and Goulding (2011) conclude that: “Although Shulman’s work was groundbreaking and his ideas continue to influence the majority of research in the field, later researchers in the same tradition argue that

it is not sufficiently developed to be operationalised in research on teacher knowledge and teacher education” (p. 12). We note that Shulman did not write specifically for mathematics teaching, but for all teaching subjects; and that his categories tend to reflect the educational context of the USA where there was no national curriculum.

36.2.2 *Michigan Model*

Ball et al. (2008), while sympathetic to Shulman, prefer to use the term Mathematical knowledge for teaching (MKT). Within this idea, they identify four constituent domains or categories: (1) Common content knowledge (CCK) defined as the mathematical knowledge and skill used in settings other than teaching; (2) Specialised content knowledge (SCK) as the mathematical knowledge and skill unique to teaching specific topics; (3) Knowledge of content and students (KCS) defined as knowledge that combines knowing about students and knowing about mathematics; and (4) Knowledge of content and teaching (KCT), which combines knowing about teaching and knowing about mathematics.

Among these four domains discussed by Ball et al. (2008), CCK is a primary component of mathematical knowledge and needs to be combined with a teacher’s SCK, the subject matter knowledge needed for teaching specific mathematics content or topics. KCS and KCT are both intended to describe distinct knowledge for teaching. However, “content” used in the four categories may refer to: today’s worksheet, or this year’s textbook, or what is contained in official curriculum documents. In this sense, KCT may not be too far removed from Shulman’s category of curriculum knowledge under which he includes teachers having a grasp of relevant materials and programs. While these knowledge domains are intended to anticipate classroom use, their instructional consequences are only implied. What is more, what appears to be a common feature of both Ball et al. (2008) and Shulman (1986, 1987) is an interpretation of “curriculum” and “curriculum knowledge” which may be based too closely on their USA experience, where curriculum knowledge can be interpreted simply as “the particular grasp of the materials and programs that serve as ‘tools of the trade’ for teachers” (Shulman, 1987, p. 8); and Ball et al. (2008) do not seem to have moved beyond this.

36.2.3 *Limitations of Research on PCK and MKT*

Ruthven (2011) has presented four distinct conceptualisations of MKT—*Subject knowledge differentiated*; *Subject knowledge contextualised*; *Subject knowledge interactivated* and *Subject knowledge mathematised*—each of which is intended to

move forward debate about and research—but in different directions. These four lines of thinking show that MKT is no longer a single unified idea. Researchers also need to be aware of the limitations of some or all of these four approaches:

- (1) All four have a strong focus on how to improve pre-service teachers' mathematical knowledge needed for their teaching in the future;
- (2) apart from the first framework adopted by Petrou and Goulding (2011), the other three do not appear to place a strong emphasis on the way in which official mathematics curriculum documents are intended to guide teaching in many countries;
- (3) apart from the first framework, the other three tend to view knowledge for teaching mathematics in general terms, rather than considering the specific areas of mathematical content important for curriculum reform;
- (4) all four conceptualisations into empirically conducted research.

Our own position on MKT is closest to that of Petrou and Goulding (2011). We use this framework to inform our construct of teacher capacity, and to show where it differs from that of Ball et al. (2008), especially in its stronger links to research on curriculum reform and school change.

36.2.4 Teacher Capacity Model

The term “Teacher capacity” comes out of the literature of school improvement, school leadership and system reform (Fullan, 2010; McDiarmid, 2006). When used in this context, teacher capacity usually relates to teachers' ability to understand and act on the reforms that policy makers are seeking to implement (Spillane, 1999). It is close to our definition of teacher capacity as professionally informed judgement and disposition to act. Researchers such as Floden, Goertz, and O'Day (1995) emphasise that teacher capacity is multidimensional and evolving. Firstly, they argue that teachers' ability to assist students in learning is dependent on teachers' own knowledge, which includes knowledge of the subject matter, knowledge of curriculum, knowledge about students and knowledge about general and subject-specific pedagogy; secondly, they argue that, while knowledge interacts with skills, there is a considerable gap between what teachers believe they should be doing in the classroom and their ability to teach in the desired ways; and thirdly, they point to the importance of dispositions, since enacting reform requires having the dispositions to meet new standards for student learning and to make the necessary changes in practice.

There are clear parallels here with Ball et al. (2008) who make the equally strong point that any definition of MKT should begin with teaching, not teachers. Any such definition must be “concerned with the tasks involved in teaching and the mathematical demands of these tasks (our emphasis). Because teaching involves showing students how to solve problems, answering students' questions and checking students' work, it demands an understanding of the content of the school curriculum” (p. 395).

36.3 Methodology

36.3.1 The Research Instrument

Teachers were invited to complete a written questionnaire consisting of two parts. Part A has four questions which were based on tasks developed in previous research, containing some situations relating to statistical thinking that students are expected to meet.

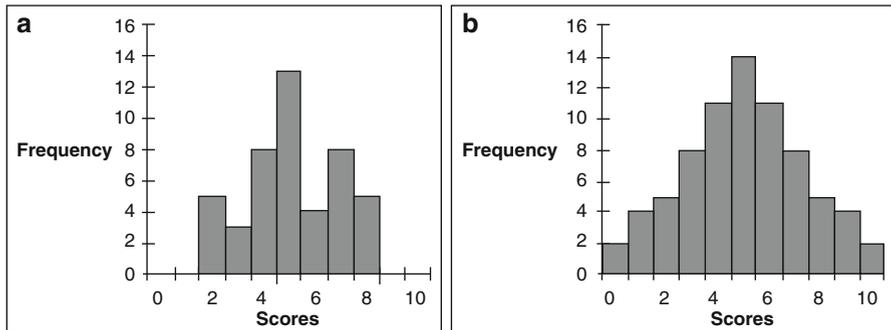
Question 1 was adapted from Shaughnessy, Canada and Ciancetta (2004):

A gumball machine has 100 gumballs in it. 20 are yellow, 30 are blue and 50 are red. The gumballs are all mixed up inside the machine.

- (a) Suppose you do the following experiment: you pick out a handful of ten gumballs, count the reds and write down the number of red gumballs in one handful. How many reds do you expect to get?
- (b) You replace the handful of ten gumballs back in the machine and mix them up again. Now you draw another handful of ten gumballs. Would you expect to get the same number of reds in every handful if you did this several times? Briefly describe why.
- (c) How many reds would surprise you in a handful of ten? Why would that surprise you?
- (d) If each time a handful of ten gumballs is taken, these are replaced and remixed before taking another handful again, what do you think is likely to occur for the numbers of red gumballs that come out for a sequence of five handfuls? Please write the number of reds in each handful here.
- (e) Look at these possibilities that some students have written down for the numbers they thought likely when they answered question d. Which one of these lists do you think best describes what is most likely to happen? Circle it. (A. 8,9,7,9,10; B. 3,7,5,8,5; C. 5,5,5,5,5; D. 2,4,3,4,3; E. 3,0,9,2,8; F. 7,7,7,7,7). Why do you think the list you chose best describes what is most likely to happen?
- (f) In the above six repetitions of the experiment, what do you think will be the highest and lowest number of reds in one handful? Please discuss briefly why you think this.

Question 2 was adapted from Meletiou and Lee (2002):

Look at the histogram of the two distributions on the right. Which of the two distributions you think has more variability? (a) Distribution A (b) Distribution B.



Briefly describe why you think this.

Question 3 was adapted from Garfield and Gal (1999):

Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80 % or more female births? (a) Hospital A (with 50 births a day); (b) Hospital B (with 10 births a day); (c) The two hospitals are equally likely to record such an event; (d) There is no basis for predicting which hospital would have that percentage of female births. Give a brief explanation of why you think like this.

Question 4 was adapted from Garfield and Gal (1999):

For 1 month, 500 elementary students kept a daily record of the hours they spent watching television. The average number of hours per week spent watching television was 28. The researchers conducting the study also obtained report cards for each of the students. They found that the students who did well in school spent less time watching television than those students who did poorly.

Which of the following statements is (are) correct? (a) The sample of 500 is too small to permit drawing conclusions; (b) If a student decreased the amount of time spent watching television, his or her performance in school would improve; (c) Even though students who did well watched less television, this doesn't necessarily mean that watching television hurts school performance; (d) One month is not a long enough period of time to estimate how many hours the students really spend watching television; (e) The research demonstrates that watching television causes poorer performance in school; (f) I don't agree with any of these statements. For one statement that you agree with, explain why you think that way. For one statement that you disagree with, explain why you think that way.

Part B of the questionnaire had three questions which asked teachers to consider teaching implications arising from the questions in Part A. Specifically, they were asked to consider common misunderstandings and difficulties for students in the Part A questions; how the key mathematical ideas or critical points presented in these questions are addressed in their respective country's official curriculum documents; and how to design some lessons to help students to understand these key ideas.

36.3.2 The Participants

There were 17 Australian secondary and primary schools randomly selected in both urban and rural regions in Melbourne. Up to four teachers of Year 6 or Year 7 in each participating school were invited to complete the questionnaire. The Australian sample consisted of 41 Australian teachers, 28 secondary teachers and 13 from primary schools. The China sample comprises 41 teachers randomly selected from training programs in Chongqing, Hangzhou and Wenzhou. Twenty eight were secondary and 13 were primary teachers.

36.3.3 Theoretical Framework

Our construct of Teacher Capacity, as professionally informed judgements and dispositions to act, is intended to capture a common ground between movements for school system and curriculum reform *and* the construct of MKT elaborated by Ball et al. (2008). Four criteria inform our theoretical model.

Criterion A: *Knowledge of Mathematics*—is intended to be applied to the tasks that the students have completed or are being asked to complete. Knowledge of Mathematics is intended to capture the key mathematical ideas for teaching specific content.

Criterion B: *Interpretation of the Intentions of the Official Mathematics Curriculum*—is concerned with how teachers relate what is mandated or recommended in official curriculum documents of China and Australia to what their students are being asked to learn. This criterion differs from MKT (Ball et al., 2008) in giving a greater emphasis to official curriculum documents and teachers' willingness to use them in planning instruction.

Criterion C: *Understanding of Students' Mathematical Thinking*—is directly concerned with teachers' capacity to interpret and differentiate between what students actually do (or did) and to anticipate what they are likely to do. It implies that teachers are able to recognise the typical errors that students make and what mathematical thinking led to these errors.

Consequently, Criterion D: *Design of Teaching*—places a clear emphasis on teachers' capacity to design teaching in order to move students' thinking forward and to respond to specific examples of students' thinking in the light of official curriculum documents. Criterion D is intended to give greater emphasis to how teachers use their professionally informed judgement to design practical teaching on specific topics.

Each criterion of above was elaborated in terms of four specific indicators (see Table 36.1).

Table 36.1 Four criteria and associated indicators

<i>Criterion A: Knowledge of Mathematics</i>	
(1)	Is the teacher able to solve the theoretical mathematical probability problem (Q1a) and be able to understand relationship between chance of real events and sample size (Q3)?
(2)	Does the teacher consistently understand the variability of theoretical probability always happens in natural events in real life (Q1b, 1d), and the variability has a certain range close to the theoretical probability (Q1c, 1e, 1f)?
(3)	Does the teacher understand the meaning of “variability” by giving specific certain information (Q2)?
(4)	Does the teacher recognise that the difference between association and causation (Q4)?
<i>Criterion B: Interpretation of the Intentions of Official Mathematics Curriculum</i>	
(1)	Does the teacher realise that “statistical thinking” should be valued in teaching and learning beyond the solutions of probability problems or does the teacher refer to relevant statements on statistical thinking in the official curriculum documents?

(continued)

Table 36.1 (continued)

(2)	Does the teacher understand and support the intention of the curriculum of helping students understand key ideas of statistical thinking such as theoretical probability, sampling, sample size and inevitable variability in actual data, rather than calculating theoretical probabilities?
(3)	Does the teacher think it important to consider statistics and probability by linking natural events and real life?
(4)	Does the teacher show in his/her descriptions of developing students' ability to read and understand data and information which is important for their further learning and future life?
<i>Criterion C: Understanding of Students' Mathematical Thinking</i>	
(1)	Is the teacher able to anticipate students' common difficulties and misconceptions on Question 1(e) in questionnaire?
(2)	Does the teacher give clear and reasonable explanations to students' incorrect answers?
(3)	Is the teacher able to discriminate between students' different levels of understanding statistics and probability according to their answers, especially discriminating between incorrect answers?
(4)	Does the teacher recognise the importance of using familiar contexts, such as coin tossing or rolling dice, to help students understand the statistical features of (less familiar) situations that contain similar statistical characteristics
<i>Criterion D: Design of Teaching</i>	
(1)	In design of teaching, does the teacher focus on the important key conceptions of statistical thinking (theoretical probability, sampling, sample size and inevitable variability in actual data, as well as using familiar contexts to simulate real world events), not focusing too much on general teaching strategies or overall descriptions on statistics and probability?
(2)	Does the teacher have the subsequent plan in next one or several lessons to respond students' incorrect answers in Question 1(e)?
(3)	Does the teacher have a long-term plan to consistently develop students' deep understanding of statistical thinking (see 1 above), not just aiming to have students correctly calculate theoretical probability problems?
(4)	Does the teacher, in his/her teaching, give concrete examples that are familiar and easy for students to help them understand statistical thinking and its relationships with real life?

36.4 Qualitative Analysis

The following examples provide evidence of Chinese and Australian teachers' (coded either as Teacher n CH or Teacher n AU) responses with respect to each of the four criteria. Teachers in both countries showed their different levels of understanding on all four criteria.

36.4.1 Criterion A (*Knowledge of Mathematics*)

Teacher 57 AU responded to Question 3 of Part A: "Hospital B is more likely to record the 80 % as it has a much smaller population... Larger samples or more trials give results that are closer to theoretical probability". This teacher clearly demonstrated understanding of the relationship between sample size and variation from theoretical probability.

However, Teacher 31 AU answered: “Both hospitals are equally likely to record 80 % female births because the probability is the same for each birth to be a boy or a girl”. This teacher considered this problem as a completely theoretical probability question and did not realise variation exists and the sample size will influence the variation. And, Teacher 28 CH didn’t identify the key point of this question by saying “it is random and no absolute result”.

36.4.2 Criterion B (Interpretation of the Intentions of Official Mathematics Curriculum)

When referred to official mathematics curriculum, Teacher 35 CH said, “This is the typical question representing thinking of probability and statistics. In the stage of primary school, statistics is more important. The main content of statistics is data processing, not to infer or guess with (theoretical) probability...”

Meanwhile, some teachers like Teacher 53 AU just listed several headings that are used in curriculum documents such as “measurement, chance and data” and some related ideas such as “calculating theoretical probabilities”. And Teacher 24 CH referred to “including mathematical thinking of abstraction, transformation, and modelling”, but could not identify any specific mathematical thinking implied in the questionnaire.

36.4.3 Criterion C (Understanding of Students’ Mathematical Thinking)

When teachers were required to comment on students’ answers in Question 1(e) of Part A, Teacher 21 AU said, “Students who choose C do not consider the variation but understand the basic principles of chance. Students who choose A, D, E and F, ... do not understand the basic principles of chance”. This teacher gave reasonable explanations for each response of students and was able to discriminate students’ different thinking level on statistics.

But these “typical” answers of students were confusing for some teachers. For example, Teacher 4 CH initially thought that “Students’ understanding is there are more red balls”, but then pointed that “Students will think all outcomes are possible, it’s difficult to judge”. This teacher did not understand the various misconceptions embedded in the “typical” answers.

36.4.4 Criterion D (Design of Teaching)

When discussed on how to help students understand the critical mathematical thinking of Question 3 of Part A, Teacher 54 AU articulated teaching plans: “There are many activities that can be carried out using counters, coins and dice to simulate certain events.

In the case of babies being born male or female, tossing a coin ten times and recording Heads as female and Tails as male could be done. If every student performs the ten tries, I would have enough data to compare and expect a good range including possibly 80 % female. I could then compare individual trials of ten to collective trials by putting together five groups of ten results and comparing the male and female numbers and hopefully show that the results tend more to 50:50 female: male”, concluding “(one) would need to get across the idea that when an experiment is conducted many times over, certain patterns are likely to appear”. This teacher correctly focused on the critical points and designed very elaborate simulation—coin tossing—which is more familiar to students, not just talking about general teaching strategies.

Teacher 19 CH indicated that the teaching focus was about “statistical knowledge”, but offered no discussion of any specific statistical concepts, giving only very general teaching strategies like having “students conduct various kinds of experiments... they need practical manipulations to explore possibility”. Likewise, Teacher 2 focused only on “understanding of fractions, percentages and decimals. I would introduce whole numbers and equivalence and converting decimals to percentages”. This teacher referred only to some general strategies like “open-ended questions including ratio of boys and girls”.

36.5 Quantitative Analysis

By assigning a score of 1 if one of the four indicators was evident in a teacher’s response, and 0 if it was omitted from their response or answered inappropriately, it was possible to construct a score of 0–4 for each criterion, and hence a maximum score of 16 across the four criteria. We allowed for the possibility that teachers might provide convincing alternative indicators to the four indicators listed.

The two researchers operated independently to score teachers’ responses; then a careful confirmative check took place in order to resolve any difference. A high degree of consistency was present in the initial grading by the two graders, where, in less than 30 cases of 0/1 grading, only minor differences occurred. Any resulting differences in grading the 82 responses across the four criteria were easily resolved by consensus.

36.5.1 *A Summary for Chinese and Australian Samples*

For the 41 Chinese teachers, the highest score was 14 and the lowest score was 3, with a median score of 8. For Australian teachers, the highest score was 15 and the lowest was 4, with a median score of 9. The respective mean scores were 8.34 (Chinese) and 9.27 (Australian) with standard deviations 2.70 and 2.63, respectively. Table 36.2 shows means and deviations that were calculated for each of the

Table 36.2 Means for each criterion and global means and deviations

Sample	Criterion A	Criterion B	Criterion C	Criterion D	Total
Chinese (41)	2.42 (0.67)	2.02 (0.96)	2.27 (0.71)	1.66 (0.97)	8.37 (2.70)
Australian (41)	2.72 (0.86)	2.21 (0.83)	2.49 (0.76)	1.85 (0.93)	9.26 (2.63)
CH and AU (82)	2.56 (0.82)	2.11 (0.90)	2.38 (0.76)	1.77 (0.97)	8.82 (2.79)

criteria and total score. Of the four criteria, Criterion D (Design of Teaching) had the lowest mean (1.77) followed by Criterion B (Interpretation of the Official Mathematics Curriculum) with the mean of 2.11, followed by Criterion C (Understanding of Students' Mathematical Thinking) with the mean of 2.38. Criterion A (Knowledge of Mathematics) had the highest mean at 2.56.

Australian teachers scored slightly higher on all four criteria than their Australian counterparts, but there was no statistically significant difference. On Criterion A, Australian teachers were clearer on the understanding critical concepts in statistics, especially in distinguishing variability from theoretical probability; on Criterion B, Chinese teachers paid more attention to methods to calculate possibility or chance that students need to learn, but Australian teachers were more focused on the development of how to deal with data in practical situations; on Criterion C, Australian teachers performed better on anticipating difficulties and misunderstandings that students might encounter; on Criterion D, Australian teachers were more likely to locate the key statistical idea in the hospital question in Part A and could show in more practical ways how to develop related statistical thinking.

36.5.2 Three Classifications of Teacher Capacity

Three subcategories of our construct of teacher capacity were created, with the boundaries set on the basis of the qualitative analysis of teachers' responses as discussed earlier. These were High capacity (score 11–16), Medium capacity (score 6–10) and Low capacity (score 0–5). These classifications using the two samples are shown in Table 36.3.

There were more High Capacity teachers in Australian sample than in Chinese sample (respectively 8 and 7); less Australian teachers were classified as Low Capacity than Chinese teachers (respectively 7 and 8). In both Chinese and Australian samples, Medium Capacity group was the biggest group which included 26 teachers out of 41, that was more than 60 %.

High Capacity teaching of statistical thinking was evident in nearly 20 % of Chinese and Australian teachers' responses to the questionnaire. It was shown by a clear understanding of the critical thinking and concepts in statistics of the four questions of Part A; relating the tasks to relevant curriculum documents; by high interpretative skills when applied to each of the six possible answers of students' work in Question 1(e); and by an extensive range of ideas for designing and implementing a teaching program to support the development of students' statistical

Table 36.3 Classifications of teacher capacity

Capacity	Chinese	Australian
High	7 (17.1 %)	8 (19.5 %)
Medium	26 (63.4 %)	26 (63.4 %)
Low	8 (19.5 %)	7 (17.1 %)

thinking. Medium Capacity was shown by approximately 60 % of teachers who, while possessing knowledge and skills supportive of these directions, clearly need to increase their current levels of professional knowledge and skills. In both samples, Low Capacity was evident in a minority of teachers—nearly 20 %—who appeared unable to express a clear articulation of the mathematical nature of the tasks, or what differentiated the six students' answers in Question 1(e). These teachers were unable to point with any confidence to how the tasks related to what is contained in official curriculum documents, and found it difficult to describe how they would plan a program of teaching to foster students' statistical ideas.

36.6 Conclusions

Our construct of teacher capacity, presented here as teachers' professionally informed judgements and dispositions to act, connects to but differs from earlier research into Pedagogical Content Knowledge by Shulman (1986, 1987) and MKT by Ball et al. (2008). Here teacher capacity was investigated in terms of Knowledge of Mathematics, Interpretation of the Intentions of Official Curriculum documents, Understanding of Students' Thinking and Design of Teaching to foster the underlying mathematical ideas. Performance on each criterion was ascertained using a precise set of indicators that were related to the specific mathematical tasks, students expected thinking in relation to those tasks, the relationship between the tasks and official curriculum documents and teachers' ability to design explicit teaching sequences to support students' learning.

Design of Teaching, informed by the other three criteria, appears to be the critical dimension for the implementation of curriculum reform; and the criterion that most clearly distinguishes between different levels of teachers' capacity to enact reform. Our construct of teacher capacity strongly reflects the view that effective implementation of any curriculum reform depends on teachers' subtle interpretations of official curriculum documents and their professional dispositions to act on those ideas, which go well beyond general descriptions or statements of intent that are usually embodied in official curriculum advice.

References

- Australian Curriculum and Reporting Authority. (2010). *The Australian curriculum: Mathematics*. Sydney: Australian Curriculum and Reporting Authority.
- Ball, D., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*, 389–407.
- Datnow, A., & Castellano, M. (2001). Teachers' responses to success for all: How beliefs, experiences and adaptations shape curriculum. *American Educational Research Journal*, *37*, 775–800.
- Department of Education and Early Childhood Development. (2006). *Mathematics developmental continuum*. Melbourne: Department of Education and Early Childhood Development.
- Floden, R., Goertz, M., & O'Day, J. (1995). Capacity building in systemic reform. *Phi Delta Kappan*, *77*(1), 19–21.
- Fullan, M. (2010). *All systems go: The change imperative for whole system reform*. Thousand Oaks, CA: Sage.
- Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning*. Milton Keynes, England: Springer.
- Garfield, J., & Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review*, *67*, 1–12.
- McDiarmid, B. (2006). *Rethinking teacher capacity*. <http://scimath.unl.edu/MIM/rew2006/Powerpoints/REW06McDiarmid.ppt>.
- Meletioui, M., & Lee, C. (2002). Teaching students the stochastic nature of statistical concepts in an introductory statistics course. *Statistical Education Research Journal*, *1*(2), 22–37.
- Ministry of Education of PRC. (2011). *Mathematics curriculum standards for compulsory education (2011 version)*. Beijing: Beijing Normal University Press.
- Petrou, M., & Goulding, M. (2011). Conceptualising teachers' mathematical knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 9–25). Dordrecht, The Netherlands: Springer.
- Ruthven, K. (2011). Conceptualising mathematical knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 83–98). Dordrecht, The Netherlands: Springer.
- Shaughnessy, J. M., Canada, D., & Ciancetta, M. (2004). Types of student reasoning on sampling tasks. In B. Philips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a Statistically Literate Society*. Voorburg, The Netherlands: International Statistical Institute.
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*, 1–22.
- Spillane, J. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies*, *31*, 143–175.
- Victorian Curriculum and Assessment Authority. (2008). *Victorian essential learning standards (Mathematics)*. Melbourne: Victorian Curriculum and Assessment Authority.

Chapter 37

A Framework for Assessing Statistical Knowledge for Teaching Based on the Identification of Conceptions of Variability Held by Teachers

Orlando González

37.1 Introduction

Aiming toward statistical literacy in today's information society, recent curricular reforms in many countries have brought a number of topics related to statistics and probability into the school mathematics curriculum (e.g., NCTM, 2000), it being noticeable that variability may arise in many different ways in such topics.

Variability—a property of an statistical object which accounts for its propensity to vary or change—is considered by several researchers as a fundamental concept in statistics (e.g., Pfannkuch & Ben-Zvi, 2011; Shaughnessy, 2007); and its acknowledgement and understanding are regarded as essential skills for statistical literacy, reasoning, and thinking (e.g., Sánchez, da Silva, & Coutinho, 2011; Wild & Pfannkuch, 1999). According to Gattuso and Ottaviani (2011, p. 122), “[t]o be part of a modern society in a competent and critical way requires citizens to ... understand the variability, dispersion, and heterogeneity which cause uncertainty in interpreting, in making decisions, and in facing risks,” and teachers are in charge to foster and develop such knowledge and skills in their students. Despite all these facts, scarce studies can be found in the literature focused on the conceptions of variability held by in-service teachers, as well as on the knowledge entailed by teaching variability-related contents, and statistics in general, to help students achieve the aims of statistics education (Shaughnessy, 2007). Hence, it is by no

O. González (✉)

Graduate School of International Development and Cooperation (IDEC),
Hiroshima University, Hiroshima, Japan
e-mail: ogonzalez@hiroshima-u.ac.jp

means surprising the urgent call for increasing research on these areas made for a number of concerned researchers, particularly for studies on teachers' professional knowledge and teachers' practices while teaching variability (e.g., Sánchez et al., 2011, p. 219), as well as for the developing and improvement of the models for the didactic knowledge required to teach statistics (e.g., Godino, Ortiz, Roa, & Wilhelmi, 2011, p. 281). Accordingly, the purpose of this chapter is to respond to such calls for research.

In the present article, I propose a conceptualization of SKT—the knowledge, skills, and habits of mind needed to carry out effectively the work of teaching statistics in a way that supports student learning and achievement—aiming to contribute to a better understanding of what knowledge is necessary and sufficient to teach statistics well, by addressing and helping to fill in some notable gaps in the research literature on statistics education. The proposed model focuses on investigating teachers' knowledge for teaching variability-related concepts, and its main implications are (a) preparing the ground for future empirical research on SKT at school level; (b) bringing closer together SKT and the model for MKT developed by Ball, Thames, and Phelps (2008); (c) extending such model to include teachers' beliefs about statistics, teaching and learning of the various statistical topics in the school mathematics curriculum; and (d) identifying and taking into account teachers' conceptions of variability.

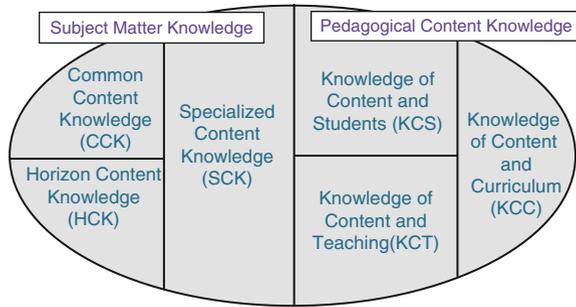
37.2 Literature Review

In the next subsections, the author presents a summary of some of the research literature relevant to the development of the framework for SKT proposed in the current article.

37.2.1 *The MKT Model*

Influenced by the criticisms directed at the aspects of teacher knowledge identified by Shulman in his breakthrough article (Shulman, 1986), and examining ways in which Shulman's ideas could be operationalized in mathematics education, Ball et al. (2008) developed the notion of mathematical knowledge for teaching (MKT), a practice-based model of content knowledge needed for teaching mathematics effectively, focused on both what teachers do as they teach mathematics and what knowledge and skills teachers need in order to be able to teach mathematics effectively. This model describes MKT as being made up of two domains—subject matter knowledge (SMK) and pedagogical content knowledge (PCK)—each of them structured in a tripartite form, as depicted in Fig. 37.1.

Fig. 37.1 Domains of MKT, according to Ball et al. (2008)



According to Ball et al. (2008), SMK can be divided into common content knowledge (CCK), specialized content knowledge (SCK), and horizon content knowledge (HCK). The construct CCK refers to the mathematical knowledge and skills expected of any well-educated adult, which are commonly used in any setting, not necessarily the one of teaching. SCK is the mathematical content knowledge specific to the work of teaching and needed in its practice—and not in the practice of other professions. The third construct, HCK, is an awareness of where both the present learner experience and the instructional content are situated over the span of mathematics included in the school curriculum, and of what their connections are with the key mathematical practices and major disciplinary ideas and structures that lie ahead, on the curricular horizon.

Furthermore, Ball and her colleagues presented a more refined division of Shulman's PCK, comprised by knowledge of content and students (KCS), knowledge of content and teaching (KCT), and knowledge of content and curriculum (KCC). The construct KCS represents the teacher's amalgamated knowledge about how students come to understand mathematics and mathematics content itself. KCT refers to the knowledge about how to carry out the design of instruction in order to develop mathematical understanding in students, and about how a particular mathematical content shapes mathematics instructional practice. Finally, KCC is the knowledge that teachers have on how specific topics, procedures, and concepts are offered in school curricula at a particular grade level, along with an understanding of the grade-wise relationships among them and the variety of educational materials that can be drawn on to facilitate the development of students' mathematical understandings.

Through this model for MKT, Ball and her colleagues made significant progress in identifying the relationship between teacher knowledge and students' achievement in mathematics, as well as in developing reliable and valid measures of MKT. Nevertheless, as highlighted by some researchers (e.g., Petrou & Goulding, 2011, p. 16), Ball et al.'s (2008) model of MKT does not acknowledge the role of beliefs in teachers' taking on, and performance of, educational practices, which could be a drawback since beliefs are often regarded in the literature as important factors affecting teachers' instructional practice.

37.2.2 *MKT-Based Models for SKT*

It is by no means surprising that almost all the few conceptualizations of SKT proposed to date have assimilated some of the categories present in the aforementioned model for MKT developed by Ball and her colleagues, due to the considerable overlap and cooperation between mathematics and statistics, as well as between the structure of mathematics education and statistics education (see Hand, 1998). Nevertheless, due to the specificity of statistics as discipline (see Gattuso & Ottaviani, 2011; Godino et al., 2011), it is not surprising either the effort that has been made through those few conceptualizations of SKT to adapt MKT components in order to meet the particular case of statistics education. In this subsection, the author presents an overview of the MKT-based models of SKT developed by Burgess (2011), Groth (2007), and Noll (2011).

Groth (2007) developed a hypothetical framework to explain the SKT required for teaching statistics at high school level, borrowing and focusing on the constructs of CCK and SCK described by Ball et al. (2008), and merging and adapting them with the framework for statistical problem solving given in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report (Franklin et al., 2007), in order to characterize the work of teaching statistics, make distinctions between the mathematical and nonmathematical knowledge needed for it, and differentiate such work from the one of teaching mathematics.

In his model, Groth argues that some aspects of the common and specialized knowledge entailed by the teaching of statistics require a growing research base, particularly the specialized one related to nonmathematical knowledge, which encompasses the pedagogical activities that take place in the classroom.

In order to examine, through a classroom-based approach, the knowledge that elementary school teachers need to successfully implement the teaching of statistics through projects and investigations, Burgess (2011) developed a two-dimensional framework comprised by four of the knowledge components described by Ball et al. (2008)—CCK, SCK, KCS, and KCT—and six out of eight components of Wild and Pfannkuch's (1999) model for statistical thinking in empirical enquiry. Through his model, Burgess identified the different types of knowledge that were either needed and used, or needed but not used, in the context of teaching experiences based on statistical investigations, finding, among other things, that all the aspects of knowledge included in his proposed model were necessary in the classroom.

Noll (2011) investigated the SKT held by 68 American graduate teaching assistants' (TAs) using a task-based survey and a series of semistructured interviews, focusing on TAs' knowledge about distributions of data and empirical samples, as well as in their knowledge of student thinking about sampling concepts. Noll selected three of the components described by Ball et al. (2008)—CCK, SCK, and KCS—to develop her framework. Key features in her model are the interpretation of CCK and SCK as statistical literacy and statistical thinking, respectively. The findings from Noll's research indicate that TAs have a limited SKT in all the three components in

study, which is particularly noticeable in their difficulty teaching certain topics—especially conceptual ideas of variability—and making sense of students’ work and interpretation about variability and other sampling-related concepts.

37.3 A New Conceptualization of SKT

The purpose of this section is to make a contribution to the literature on statistics education, by proposing a conceptualization attempting to characterize some critical components related to the knowledge required to teach statistics effectively. On the basis of literature review and personal research experience, several components that were thought to be potential predictors of SKT were identified and considered for analysis, and the following arguments were raised as a result of such analysis:

- (a) *The proposed model of SKT should be closely tied to a model of MKT:* On the basis that school statistics is often taught as part of mathematics curriculum by mathematics teachers, as well as due to the common grounds shared by mathematics and statistics, it is anticipated that a model of SKT should be closely tied to a model of MKT. Consequently, I argued that the six constructs necessary for having a solid MKT identified by Ball et al. (2008) in their framework would serve as a useful starting point to hypothesize what knowledge might be needed for teaching statistics effectively.
- (b) *Some knowledge components in the MKT model used must be redefined to meet the requirements of teaching statistics:* Although mathematics and statistics share some common grounds, the two disciplines are different in several ways (an in-depth discussion of these differences can be found in Gattuso & Ottaviani, 2011). Therefore, in order to acknowledge such differences and meet the requirements specific to the teaching of statistics, some knowledge components in the MKT model used must be redefined. In the case of the conceptualization proposed here, CCK will be seen as statistical literacy, which development is regarded as one of the main goals of statistics education and mathematics curricula at all educational levels (e.g., Gal, 2004; Pfannkuch & Ben-Zvi, 2011), and thus its acquisition is expected from any individual after completing school education. The rest of knowledge components in this framework are defined in the same way as in the model of MKT by Ball et al. (2008), but rephrased in some cases to meet the requirements of teaching statistics.
- (c) *In order to conceptualize SKT, teachers’ beliefs about statistics, teaching and learning must be considered:* The relationship between beliefs—defined by Philipp (2007, p. 259) as “psychologically held understandings, premises, or prepositions about the world that are thought to be true”—and teachers’ classroom practice has been well articulated in the literature by several researchers (e.g., Gal, Ginsburg, & Schau, 1997; Philipp, 2007; Pierce & Chick, 2011). Moreover, beliefs are identified by Gal (2004) as one of the dispositional elements of statistical literacy, being the latter regarded as CCK in the present conceptualization of SKT. On the basis of these facts, in this model of SKT

teachers' beliefs about statistics, teaching and learning are going to be regarded as fundamental, attempting in that way to obtain a much richer and broader picture of the knowledge needed to teach statistics efficiently, as well as to overcome a common drawback in all the MKT-based frameworks of SKT reviewed previously.

- (d) *Tasks designed to elicit teachers' conceptions of variability would be helpful to provide indicators to measure SKT as defined in this study:* In the case of teachers, conceptions—the set of internal representations and corresponding associations that a mathematical concept evokes in the individual—have been proved to influence their own approaches to teaching, and consequently their students' approaches to learning (e.g., Trigwell, Prosser, & Waterhouse, 1999). Also, the work carried out by González (2011) and Isoda and González (2012) provides empirical evidence that the use of tasks addressing variability and variability-related concepts is an effective method for eliciting, identifying, describing, and assessing not only the conceptions of variability held by teachers but also their SMK in statistics. On the basis of these facts, and because conceptions represent knowledge and beliefs working in tandem (Knuth, 2002), gaining insight into the teachers' conceptions of variability is regarded as necessary in the proposed model for SKT.

37.3.1 An Instrument to Assess SKT Based on the Identification of Conceptions of Variability Held by Teachers

Based on the four arguments outlined above, a pen-and-paper instrument, comprised by tasks addressing variability and variability-related concepts present in the school mathematics curriculum, was designed in order to assess the eight components of SKT identified and described by this study—the six knowledge components in the model for MKT developed by Ball et al. (2008); teachers' beliefs about statistics, teaching and learning; and teachers' conceptions of variability. Each item in the instrument was developed based on questions used in previous studies with similar aims reported in the literature (e.g., Ball et al., 2008; Isoda & González, 2012), which were adapted to reflect the context of the item, the case of teaching school statistics, and the specific objectives of the present conceptualization of SKT.

In order to provide a comprehensive framework for conceptualizing SKT in the context of variability, twelve indicators were identified and selected for assessing SKT from the teachers' answers to each of the designed items (see Table 37.1).

Item 1 is provided as an example of the designed items (see Fig. 37.2). The original version of the task (by Garfield, delMas, & Chance, 1999) was adapted and enriched with questions aiming to elicit all the facets of SKT identified by this framework. A mapping between the components of SKT that would be brought out by each question in Item 1 and the indicators associated to such components identified by this framework could be appreciated in Table 37.2.

Table 37.1 Set of proposed indicators to assess SKT through the answers to Item 1

A: Indicators associated to Statistical Literacy (CCK)

1. Is the teacher able to give an appropriate and correct answer to the given task?
2. Does the teacher consistently identify and acknowledge variability and correctly interpret its meaning in the context of the given task?

B: Indicators associated to SKC

1. Does the teacher show evidence of ability to determine the accuracy of common and nonstandard arguments, methods, and solutions that could be provided on a single question/task by students (especially while recognizing whether a student's answer is right or not)?
2. Does the teacher show evidence of ability to analyze right and wrong solutions that could be given by students, by providing explanations about what reasoning and/or mathematical/statistical steps likely produced such responses, and why, in a clear, accurate and appropriate way?

C: Indicators associated to HCK

1. Does the teacher show evidence of having ability to identify whether a student comment or response is mathematically/statistically interesting or significant?
2. Is the teacher able to identify the mathematically/statistically significant notions that underlie and overlie the statistical ideas involved in the given task?

D: Indicators associated to KCS

1. Is the teacher able to anticipate students' common responses, difficulties, and misconceptions on the given task?
2. Does the teacher show evidence of knowing the most likely reasons for students' responses, misconceptions, and difficulties in relation to the statistical ideas involved in the given task?

E: Indicators associated to KCT

1. In design of teaching, does the teacher show evidence of knowing what tasks, activities, and strategies could be used to set up a productive whole-class discussion aimed at developing students' deep understanding of the key statistical ideas involved in the given task, instead of focusing just on in computation methods or general calculation techniques?
2. Does the teacher show evidence of knowing how to sequence such tasks, activities, and strategies, in order to develop students' deep understanding of the key statistical ideas involved in the given task?

F: Indicators associated to KCC

1. Does the teacher show evidence of knowing at what grade levels and content areas students are typically taught about the statistical ideas involved in the given task?
2. Does the designed lesson (or series of lessons) show evidence of teacher's understanding and support of the educational goals and the intentions of the official curriculum documents in relation to the teaching of the statistical contents present in the given problem, as well as statistics in general?

The context of the task posed in Item 1—comparing distributions—has been acknowledged as “a fruitful arena for expanding teachers’ understanding of distribution and conceptions of variability” (Makar & Confrey, 2004, p. 371). Moreover, giving an appropriate answer to Item 1 requires from teachers, among others things, knowledge and understanding of several fundamental concepts and ideas in school statistics—as in Questions (a) and (b)—; ability to connect and represent such concepts and ideas—as in Questions (a), (b) and (e)—; ability to make sense of students’ answers and to sort out the reasonable ones from those that are incorrect—as in

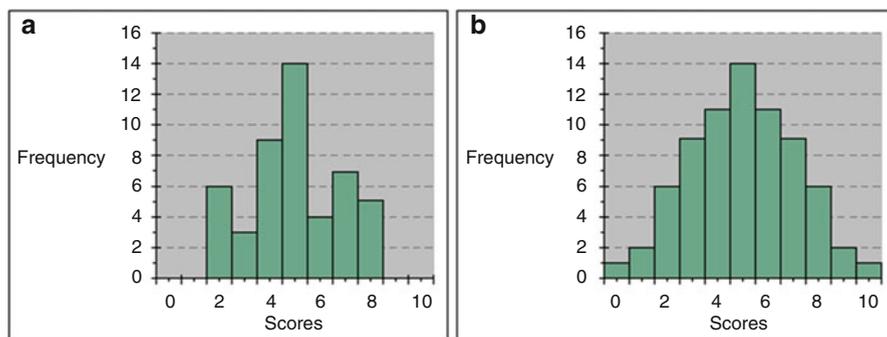


Fig. 37.2 Item 1 — “Choosing the distribution with more variability” task

Table 37.2 Knowledge components of SKT elicited by each of the questions posed in Item 1

Elicited knowledge component of SKT	Associated indicator of SKT	Question
Statistical literacy (as CCK)	A1	(b)
	A2	(b)
Specialized content knowledge (SCK)	B1	(c)
	B2	(c)
Horizon content knowledge (HCK)	C1	(e)
	C2	(a)
Knowledge of content and students (KCS)	D1	(d)
	D2	(d)
Knowledge of content and teaching (KCT)	E1	(g)
	E2	(g)
Knowledge of content and curriculum (KCC)	F1	(f)
	F2	(g)

Question (c)—; understanding of how students reason in the context of the given task—as in Question (d)—; knowledge about how the concepts and ideas involved in the posed task are developed curriculum-wise as one move up the education ladder—as in Question (f)—; knowledge about how to interpret and teach different, but interconnected and interdependent, variability-related concepts used in statistics, as well as how to teach and put into practice the statistical habits of mind related to them—as in Question (g). Also, since teachers are expected to know how to map the characteristics of the given histograms to alternate representations in order to provide an evidence-based statistical argument to justify and defend their answers (see Fig. 37.3), it is anticipated that Question (g) will also elicit how teachers promote the development of statistical discourse and argumentation into the classroom, which is of crucial importance to develop statistical literacy and avoid students’ misperceptions of statistics (Gal, 2004; Pfannkuch & Ben-Zvi, 2011), as well as to make visible teachers’ ability to recognize what concepts can be addressed through a particular data set, and to plan and implement effective learning in the

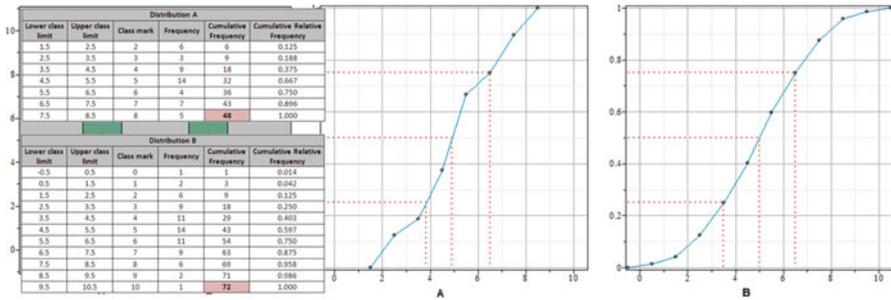


Fig. 37.3 Frequency distribution tables, boxplots, and ogives are some of the alternate representations and connections expected from teachers when dealing with Item 1

classroom with data, abilities required from teachers in order to be competent in developing statistical literacy in their students (Batanero & Díaz, 2010). Based on these arguments, all of them strongly related to specific components identified in the model of SKT proposed by this chapter, the selection of the task and questions posed in Item 1 is amply justified.

Regarding teachers’ beliefs, a useful way to identify them is from how teachers answer to students’ thinking in the classroom, interpret official curriculum documents, and design learning activities. For example, by answering to Question (g) in Item 1, it is anticipated that teachers’ personal approaches to teach specific statistical contents will give evidence of their beliefs about whether, for example, teaching and learning of statistics is better accomplished through emphasizing the memorization of formulas and procedures, rather than through the developing of students’ conceptual understanding of statistics and their ability to apply and interpret statistics in meaningful ways (Pierce & Chick, 2011).

Finally, regarding teachers’ conceptions of variability, it is anticipated that several characteristics about how teachers acknowledge and describe variability in the context of the given task will emerge through their answers to Question (b) in Item 1. The types of conceptions of variability identified by Shaughnessy (2007, pp. 984–985) will be used in this study to classify those distinguished in teachers’ answers through the proposed framework.

37.4 Conclusions

To teach statistics at any educational level, teachers must grapple with the concept of variability, one with which students often struggle. In this article, it is argued that teachers’ SMK and PCK (as in Ball et al., 2008), beliefs, and conceptions of variability (as in Shaughnessy, 2007) play altogether an important role in the shaping and effectiveness of the teaching practice in statistics. Therefore, after a literature review and theoretical considerations, a model for SKT combining the aforementioned facets was developed, and an approach for the upcoming empirical research was presented.

The conceptualization of SKT being arisen in this article not only attempts to respond to the calls that have been made for more research on particular issues in statistics education, but also proposes posing tasks that involve dealing with variability as a way to assess specific components of the knowledge needed by teachers to teach school statistic effectively.

Since new school mathematics curricula worldwide require from teachers competence to build and scaffold students' statistical knowledge and conceptions, and to help their students to develop both their ability to think and reason statistically and their statistical argumentation (Pfanckuch & Ben-Zvi, 2011), answers to Item 1 are anticipated to provide enough information on how developed these knowledge and skills are in our school mathematics teachers.

Finally, the proposed conceptual framework attempts to serve as a useful tool for discussing about in-service teachers' knowledge and skills in contexts in which variability may arise, and upcoming empirical research using this framework, as well as continued work in this area, may bring about further refinements to the conceptualization proposed.

References

- Ball, D. L., Thames, M., & 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.
- Burgess, T. A. (2011). Teacher knowledge of and for statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 259–270). New York: Springer.
- 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.
- Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht, The Netherlands: Kluwer.
- Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37–51). Amsterdam, The Netherlands: Ios Press.
- Garfield, J., delMas, R., & Chance, B. L. (1999). *Tools for Teaching and Assessing Statistical Inference*. Retrieved from http://www.tc.umn.edu/~delma001/stat_tools/
- 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: Springer.
- Godino, J. D., Ortiz, J. J., Roa, R., & Wilhelmi, M. R. (2011). Models for statistical pedagogical knowledge. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 271–282). New York: Springer.
- González, O. (2011). *Survey on Japanese elementary, middle and high school teachers' statistical literacy: Focusing on variability* (Unpublished master's thesis). University of Tsukuba, Tsukuba.

- Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, 38, 427–437.
- Hand, D. J. (1998). Breaking misconceptions—Statistics and its relationship to mathematics. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47, 245–250.
- Isoda, M., & González, O. (2012). 小学校・中学校・高等学校教師の統計的リテラシーに関する調査研究: 「ばらつき(あらばり)」に係る教員研修の必要性を探る [Survey on elementary, junior and high school teachers' statistical literacy: The need for teacher training in variability]. *Journal of Science Education in Japan*, 36(1), 61–76.
- Knuth, E. (2002). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of Mathematics Teacher Education*, 5(1), 61–88.
- Makar, K. M., & Confrey, J. (2004). Secondary teachers' reasoning about comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenges of developing statistical literacy, reasoning, and thinking* (pp. 327–352). Dordrecht, The Netherlands: Kluwer.
- National Council of Teachers of Mathematics (NCTM) (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Noll, J. A. (2011). Graduate teaching assistants' statistical content knowledge of sampling. *Statistics Education Research Journal*, 10(2), 48–74.
- Petrou, M., & Goulding, M. (2011). Conceptualising teachers' mathematical knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (pp. 9–25). New York: Springer.
- Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 323–333). New York: Springer.
- Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257–315). Charlotte, NC: Information Age.
- Pierce, R., & Chick, H. (2011). Teachers' beliefs about statistics education. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 151–162). New York: Springer.
- Sánchez, E., da 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). New York: Springer.
- Shaughnessy, J. M. (2007). Research on statistics' reasoning and learning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Reston, VA: NCTM.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, 37, 57–70.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–265.

Chapter 38

An Investigation into the Statistics Education of Preservice Mathematics Teachers in an Irish University

Olivia Fitzmaurice, Ailish Hannigan, and Aisling Mary Leavy

This chapter investigates how well second-level preservice mathematics teachers at one Irish third-level teacher education college are prepared to teach statistics. An empirical study on the conceptual understanding of core statistical concepts of these preservice teachers as they prepare to graduate is presented. This issue is pertinent internationally (Shaughnessy, 1992; Zieffler et al., 2008) but particularly in Ireland currently as a new mathematics curriculum at second level, entitled “Project Maths,” was rolled out to all schools nationally in September, 2010. This new curriculum included a move from a situation where Statistics and Probability was previously an optional component of the exit-level school state examinations to a situation where it forms a compulsory one-fifth of the new curriculum. A study was carried out on 115 (86 % of the total 134) preservice mathematics teachers in the institution to assess how well prepared these preservice teachers are to teach statistics when they qualify. The chapter concludes with a discussion of the implications of these findings and the potential impact on their future students.

O. Fitzmaurice (✉) • A. Hannigan
Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland
e-mail: olivia.gill@ul.ie; ailiah.hannigan@ul.ie

A.M. Leavy
Immaculate College, Immaculata, PA, USA
e-mail: aisling.leavy@mic.ul.ie

References

- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematical teaching and learning* (pp. 465–494). New York: McMillan.
- Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education*, 16(2), www.amstat.org/publications/jse/v16n2/zieffler.html.

Index

B

- Bootstrapping confidence intervals
 - big ideas, 187–188
 - statistical inference, 182
 - in statistics education, 183–184
 - task description, 184–186
 - two research cycles, 184
 - variation, 182
 - verbalizations, 188–189
 - visualizations, 186–187
- Bottom-up analysis, 86
- Box plot, 125
 - data corpus and analysis, 241–242
 - data set, 242–244
 - definition, 235
 - features, 236, 238, 244
 - pretest scenario, 236
 - quartile/interquartile range, 238
 - second and third quartiles, 238
 - Tinkerplots software
 - definition, 236
 - instructional method and sample activities, 238–241
 - prior statistical skills and knowledge, 236
- 2-By-2 contingency tables, 77–79

C

- Camtasia*TM, 28
- Camtasia relay recordings, 224
- Central statistical organization (CSO), 228
- Classroom questions
 - analysis, 275
 - approaches, 275
 - classification, 275

- close questions, 274
- Colombian curriculum, 273
- display, 273
- high-order questions, 272
- implications, 277
- indirect methods, 272
- mathematics curriculum and education, 273
- methodology, 273–274
- monitoring, 274
- percentage of, 275
- procedural, 274
- referential, 273
- in teaching and learning, 272
- types of, 272
- Common content knowledge (CCK), 303, 317
- Contextual and statistical spheres, 37–38
- Correlation
 - 2-by-2 contingency tables, 77–79
 - contingency tables, 72
 - Pearson correlation coefficient, 71
 - research methods, 73
 - 2007 Revised National Curriculum, 72
 - scatter plots, 71, 73–77
 - tools of presentation, 73
- Critical incidents (CIs), 252
- Critical thinking and statistical thinking (CTST)
 - advantages and disadvantages, 85
 - appropriate model and statistical variation, 90
 - computer magazine study, 89
 - empirical study, 85
 - evaluation, 89
 - inappropriate statistical model, 88
 - manipulation data, 84, 85
 - mathematical/statistical model, 88
 - methods, 86

- Critical thinking and statistical thinking (*cont.*)
 population development, 87
 questioning data, 91
 rational thinking processes and decisions, 85
 research question, 85
 speedy recovery, 91
 statistical literacy, definition of, 84
 statistical value, 84
 traditional and extensive descriptions, 86
- D**
- Data and context
 after school activities, 29
 building connections, 28
 data analysis, 29
 EDA, 25
 experienced researchers, 26
 gender activity, 31
 growing samples, 27
 IIR, 27
 integraton, 33
 interpretation of, 26
 investigation cycles, 30
 ISI, 26
 method, 28
 process development, 28
 setting and participants, 28
 statistical investigation, 26
- Data distributions, 117
- Developing statistical literacy (DSL), 299, 300
- Didactical sequences (DS), 125
 AVALE-EB, 281, 284, 285
 computing environment, 282, 284
 data, 284
 pencil and paper environment, 282, 283
 profile class, 281, 283
 scientific investigation, 281
 treatment of information, 281
 uses, 282
 watching games, 283
 Water Planet, 281, 284
- Distribution description framework (DDF),
 59, 60
- Distribution shape
 DDF, 59
 definition, 52
 dot plots and stem-and-leaf plots, 52
 pre- and post-test written responses,
 57–58
 pyramid, 53
 research method, 53, 54
 semi-circle and bell shaped, 53
 single univariate distributions, 52
- teaching activities
 data graphs, 57
 linking shape and context, 55–57
 seeing and describing shape, 55
 theoretical frameworks, 53
- Dot plot, 125
- Dynamic software, 183
- E**
- Illuminate software, 221
- Event history calendars (EHCs), 252
- Exploratory data analysis (EDA), 25
- F**
- Fathom®, 140
- Fischbein's theory, 194
- G**
- Graph theory, 231
- Guidelines for Assessment and Instruction in
 Statistics Education (GAISE)
 Report, 318
- H**
- Hats averaging problem, 123
- Histogram, 120
- Horizon content knowledge (HCK), 317
- Human resources development, 205
- Hypothetical learning trajectory, 117
- I**
- Indian Council of Agricultural Research
 (ICAR), 230
- Indian statistical system. *See* Indian statistics
 education
- Indian statistics education
 agricultural, 229
 computer-based graphics and
 methods, 230
 CSO, 228
 improvement strategy, 231–232
 NSSO, 228
 statistical methods, 230
- Informal inferences, 27
- Informal inferential reasoning (IIR), 5, 25, 27,
 38, 140
- Informal language, 152
- Informal statistical inference (ISI), 25, 26
- Intuitive confidence interval, 44

K

- Knowledge of content and curriculum (KCC), 317
- Knowledge of content and students (KCS), 303, 317
- Knowledge of content and teaching (KCT), 303, 317

L

- Lexical ambiguity, 151, 152
- Local instruction theory, 117

M

- Mathematical knowledge for teaching (MKT) model
 - CCK, 317, 318
 - HCK, 317
 - KCC, 317
 - KCS, 317, 318
 - KCT, 317, 318
 - limitations, 303–304
 - Michigan model, 303
 - SCK, 317, 318
- Measurement error, 119
- Mental models, 15
 - definition, 15
- Michigan model, 303
- Middle school mathematics course teaching
 - statistics, 121
- Ministry of Agriculture (MOA), 228, 229
- Ministry of Education (MOE), 177, 205
- Modeling and simulation approach
 - activity guided students, 144
 - card task, 144
 - CATALST curriculum, 141, 142
 - challenges, 146, 148
 - Facebook and Music tasks, 144
 - Fathom®, 140
 - Mode level, 144
 - MOST assessment, 141, 142
 - pedagogical value, 141
 - phases, 142, 145
 - representativeness, 140
 - statisticians and educators, 139
 - TinkerPlots™, 140, 142
 - variability, 140
- Monte Carlo method, 183
- MyStatLab, 224

N

- National Sample Survey organization (NSSO), 228

O

- Online courses
 - computer/video screen capture, 220
 - content-building software, 218–220
 - demographics, 216
 - evaluation, 223
 - GAISE guidelines, 217
 - grading, 223
 - hardware, 218–220
 - interactive communication programs, 220–222
 - setup, 216
 - technology and recommendations
 - assessment, 224

P

- Pedagogical content knowledge (PCK), 316
 - limitations, 303–304
 - Shulman's model, 302
- Pirie–Kieren dynamical model, 123
- Preservice mathematics teachers, 327
- Pre-service teachers
 - dispersion, measure of, 285
- DS
 - AVALE-EB, 281, 284, 285
 - computing environment, 282, 284
 - data, 284
 - pencil and paper environment, 282, 283
 - profile class, 281, 283
 - scientific investigation, 281
 - treatment of information, 281
 - uses, 282
 - watching games, 283
 - Water Planet, 281, 284
- principles, 281
- statistical test, 285
- Super Trunfo®, 283
- technical and conceptual knowledge, 281
- theory of probability, 280
- Preservice teachers difficulties
 - connecting terms conceptually, 265–267
 - factors, 267
 - improving meta-cognitive ability, 262
 - problem-solving strategies, 268–269
 - problem-solving tasks, 262
 - writing assessment, 263–265
 - writing task, 262
- Primary teachers
 - center and spread interviewer, 295
 - curricular knowledge, 290
 - data collection, 291–292
 - limitations, 292
 - line plot, 293, 294

- Primary teachers (*cont.*)
 pedagogical matter knowledge, 290
 quantitative and qualitative techniques, 291
 research question, 291
 setting and participants, 291
 statistical knowledge, 290
 stem-and-leaf plot, 293, 294
 subject matter knowledge, 290
 ProctorU, 223
- Q**
 Q&A Discussion Board (QADB), 222
- R**
 2007 Revised National Mathematics Curriculum, 72
- S**
 Sampling
 distribution, 39, 44–45
 informal inferential reasoning, 38
 intuitive confidence interval, 39, 44
 law of large numbers, 39
 vs. population, 45
 quantitative impression, 42
 random process, 39, 44
 research question, 40
 size, 39, 42
 situations, 117
 statistics reports, 40
 tasks, 40–42, 46–47
 Sampling distribution
 data collection and analysis, 64
 NE03's response to Q1b, 66
 NM12's responses to Q2a, 66
 participants, 63
 tasks, 64
 TE16 response to Q2a, 68
 TE16 response to Q2b, 68
 TE23's responses to Q2a and Q2b, 66
 TE31 response to Q2a and Q2b, 67
 TM20-Q2a, 69
 Shulman's model, 302–303
 Small Group Discussion Board (SGDB), 222
 SoftChalk, 218, 219, 224
 SOLO classification method, 103
 SOLO model, 64
 Specialised content knowledge (SCK), 303, 317
 StatCrunch, 218, 219, 224
- Statistical information
 classification, 103–104
 critical evaluations
 factors analysis, 107
 gender difference, 105
 statistic teaching, objective of, 109
 students level, 105
 students' participation, 106–107
 test questions, 112–114
 limitations, 112
 questionnaire preparation, 102–103
 sample selection, 102
 statistical conception, 101
 teaching measures
 complete statistical survey activity, 111
 situations and diagrams, 110, 111
 skills and methods, 109–110
 Statistical investigations, 119, 120
 Statistical knowledge for teaching (SKT) conceptualization
 knowledge components, 319
 proposed model, 319
 teachers' beliefs, 320
 teaching and learning, 320
 variability, 320–323
 MKT model
 CCK, 317, 318
 HCK, 317
 hypothetical framework, 318
 KCC, 317
 KCS, 317, 318
 KCT, 317, 318
 SCK, 317, 318
 variability, 315
 Statistical symbols
 Cartesian system, 164
 choice of symbols, 167–168
 in learning statistics, 170–172
 methods, 165–166
 object/representation pairs, 164, 165
 Onto-semiotic research, 164
 rationale, 165
 research questions, 163–164
 right and left-hand symbols, 164
 standard deviation, 168–169
 variable, 170
 Statistical terms
 Byeon-In, 154
 coding, 155–159
 Bun-San, 155
 Byeon-In, 155–156, 159
 Do-Su, 156, 157, 160

- Gye-Geup, 157, 160
 - Pyo-Bon, 158, 159
 - in secondary education, 153
 - Statistical thinking, 122
 - Statistical variation
 - comfortable learning environment, 253
 - critical reflection, 250–251
 - data-centric perspective, 250, 255–256
 - design perspective, 250, 254–255
 - modeling perspective, 250, 257–258
 - phenomenological methods
 - data analysis, 252, 253
 - data and instruments, 251–252
 - policy makers and educators, 249
 - rational discourse, 251
 - teacher quality, 249
 - transformative learning theory, 250
 - trigger dilemmas, 251, 253
 - Statistics
 - anxiety, 2
 - content, 98
 - EDA and probability, integration of, 5
 - education research, 6
 - International perspective, 6–7
 - learning, 2–3
 - ongoing empirical effort, 5
 - studies, number of, 6
 - teaching, 4
 - technology, 6
 - thinking, 4, 8
 - TSG-12, 7
 - Worldwide, 2
 - Statistics education, KNOU mobile learning system
 - cyber universities conferring, 207
 - delivery system, 207
 - developing courseware and supplementary learning resources, 208
 - e-learning adaptation, 209
 - e-learning course management plan, 208
 - e-learning system, 206
 - flexible teaching–learning system, 208
 - full-scale e-learning project, 208
 - high and dynamic demand, 207
 - mobile campus, 211–212
 - MOE, 205
 - readiness, 209–210
 - standardization and quality control process, 210
 - ubiquitous learning campus, 211
 - Students misconceptions, 119
 - Subject matter knowledge (SMK), 316
- T**
- Tables, treatment of, 95
 - Teacher capacity
 - MKT
 - limitations, 303–304
 - Michigan model, 303
 - model, 304
 - participants, 306
 - PCK
 - limitations, 303–304
 - Shulman’s model, 302–303
 - qualitative analysis
 - design of teaching, 309–310
 - knowledge of mathematics, 308–309
 - official mathematics curriculum, 309
 - students’ mathematical thinking, 309
 - quantitative analysis
 - Australian teachers, 311
 - Chinese teachers, 310
 - high capacity, 311
 - low capacity, 312
 - medium capacity, 312
 - research instrument, 305–306
 - theoretical framework, 307
 - Textbook analysis, 97
 - TinkerPlots™ (TP), 140
 - Tinkerplots software
 - definition, 236
 - instructional method and sample activities, 238–241
 - prior statistical skills and knowledge, 236
 - Topic Study Group 12 (TSG-12), 7
 - Transformative learning theory, 250
- U**
- Undergraduate students conceptions
 - dynamic sketch, 195–196
 - interview protocol, 196–197
 - mean and standard deviation
 - description, 197
 - in dynamic mathematics environments, 198–202
 - statistical variability, 193
 - theoretical framework, 194–195
- V**
- Variability, 117
 - Variation, conceptions of, 175, 176
 - Virtual Environment for Supporting Statistics Literacy in Basic Education (AVALE–EB), 125, 281

Y

Young students mental modelling

contemporary period, 13

method

data, 18

material and procedure, 18

participants, 18

Piagetian period, 13

post-Piagetian period, 13

theoretical framework

essential process, 15

qualitative results, 21–22

qualitative simulations, 15

quantitative results,

19–20

thinking development,

14–15

theory based task construction

car task, 17

complexity levels, 16

money game task, 16, 17

statistical situation, 13