INVESTIGATING THE TEACHING OF STATISTICS WITH TECHNOLOGY AT THE HIGH SCHOOL LEVEL THROUGH THE USE OF ANNOTATED LESSON PLANS

by

Elizabeth Grace Arnold

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Mathematics

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2016

©COPYRIGHT

by

Elizabeth Grace Arnold

2016

All Rights Reserved

DEDICATION

This dissertation is dedicated to my wonderful parents, David and Mary Arnold. Thank you both for all of your support and encouragement throughout my entire educational journey. I love you!

ACKNOWLEDGEMENTS

I would not have been able to complete this dissertation without the support, encouragement, advice, and friendship of many people. First and foremost, I would like to express my sincere gratitude to my committee members: Dr. Elizabeth Burroughs, Dr. Megan Higgs, Dr. Jennifer Green, Dr. Mary Alice Carlson, and Dr. Nicholas Lux. I appreciate your endless support, driving me to challenge myself in ways that I did not think I could. It has been an honor to work with all of you and learn from you. I offer a special thank you to Beth, who has worked with me since I was an undergraduate student in her Linear Algebra course. You have shown me what a great teacher and advisor looks like. Thank you for being my advisor, colleague, and friend.

I am also grateful for my colleagues and professors at Montana State University. I have thoroughly enjoyed the close knit group of mathematics educators I have had the privilege of working with. Elizabeth, I will always think of the term "procrastibaking" when I should be working but am baking instead. I would also like to thank Katie Banner for her constant encouragement, friendship, and input, particularly during the last year of our dissertations. We did it! Finally, I am indebted to the five teachers who participated in my study – thank you for giving your time to help me with my research.

I would also like to thank my family who have constantly supported me. Dad, I will always enjoy the times we worked on mathematics and statistics together. Lastly, to Adam "Golden Eagle," I could not have finished this dissertation without your support and understanding. I look forward to our future together (with the pugs).

TABLE OF CONTENTS

1. THE PROBLEM 1

Introduction 1
Background 3
Statement of the Problem 11
Research Questions 13
Importance of the Study 14
List of Definitions and Acronyms 16

2. REVIEW OF THE LITERATURE 18

Introduction 18
History of K-12 Statistics Education 19
Curriculum and Evaluation Standards for School Mathematics 20
Principles and Standards for School Mathematics 22
Guidelines for Assessment and Instruction in Statistics Education 23
Common Core State Standards for Mathematics 25
The Mathematical Education of Teachers 27
The Statistical Education of Teachers 29
Statistics versus Mathematics 31
Statistical Literacy, Reasoning, and Thinking 35
Statistical Literacy 37
Statistical Reasoning 38
Statistical Thinking 40
The Relationship Between Statistical Literacy, Reasoning, and Thinking 42
Statistical Literacy in the *Pre-K-12 GAISE Report* 43
The Teaching of Statistics 44
The Culture of K-12 Statistics 46
K-12 Teachers' Statistical Content Knowledge 47
The Role of Technology 50
Knowledge and Teaching 56
Technological Pedagogical Content Knowledge 57
The Components of TPACK 58
TPACK Framework in Mathematics 60
TPACK Framework in Statistics 62
Measuring TPACK 64
Improving Teaching 67
Theoretical Framework 70

TABLE OF CONTENTS – CONTINUED

Summary 72

3. METHODOLOGY 74

Introduction 74
The Researcher 76
The Research Design 78
Population of Interest 80
The Statistics Units 82
Participants 86
Recruiting and Selecting Participants 86
Treatment Allocation 88
TPACK Rubric 96
Implementation of the Statistics Lessons 101
Post-Implementation of the Statistics Lessons 102
Data Collection 102
Surveys 103
Enrollment Survey 103
Initial Survey 103
Exit Survey 107
Teacher Background and Experience Survey 107Ta
Classroom Observations 108
Interviews 110
Summary of Data Collection 111
Analyzing the Data 112
Pre-Experiment Data Analysis 113
Post-Experiment Data Analysis 121
Classroom Observations 122
Initial and Exit Surveys 136
Interviews 137
Teacher Background and Experience Survey 137
Quality and Rigor of the Research Study 138

4. RESULTS 140

Introduction 140
Final Group of Participants 141
Background Information for the Final Five Participants 143
Quantitative Results 145
Graphical Displays of the Data 146

TABLE OF CONTENTS – CONTINUED

Investigating Research Question #1 149
Investigating Research Question #2 153
Investigating Research Question #3 159
Summary 166
Describing the Two Individual Cases 168
Karen 169
Robin 169
Karen and Robin's Initial Surveys 171
Karen's Awareness of the Use of Inquiry 172
Robin's Awareness of the Use of Inquiry 173
Karen's Awareness of the Use of Technology 175
Robin's Awareness of the Use of Technology 176
Lesson 3: Observational Studies and Experiments 178
Implementation 180
Karen's Implementation 181
Robin's Implementation 184
Karen and Robin's Technology Scores 188
Analysis of Implementation 189
Technology 190
Questioning Techniques 198
Information Giving 200
Pacing of the Lesson 204
Content 207
Lesson 4: Interval Estimation and Margin of Error 209
Implementation 211
Karen's Implementation 211
Robin's Implementation 212
Karen and Robin's Technology Scores 213
Analysis of Implementation 214
Technology 214
Questioning Techniques 223
Information Giving 226
Pacing of the Lesson 228
Content 230
Overall Analysis of Karen and Robin's Implementations 230
Karen and Robin's Exit Surveys 238
Awareness of the Use of Inquiry 238
Awareness of the Use of Technology 241
Overall Summary 243

TABLE OF CONTENTS – CONTINUED

5. CONCLUSIONS 247

Introduction 247
Summary of Key Results 248
Type of Statistics Curriculum 249
Implications 250
Implications for Teacher Educators 250
Implications for Researchers 254
Improving Teaching versus Improving Teachers 258
Future Research 259
Limitations 262
Concluding Statements 263

REFERENCES CITED 266

APPENDICES 276

APPENDIX A: Annotated Statistics Unit 277
APPENDIX B: Consent Form 361
APPENDIX C: Adjusted TPACK Rubric 363
APPENDIX D: Surveys 367
APPENDIX E: Observation Protocol 378

LIST OF TABLES

Table Page

1. A list and description of the statistics content standards recommended to be taught in an "*Inferences and Conclusions from Data"* unit within Algebra II and Mathematics III (NGA, 2010b) 82

2. The topic(s) of each lesson within the two statistics units....83

3. Comparing the language between the top two levels (exploring and advancing) of the original TPACK rubric within the curriculum component (Tournaki & Lyublinskaya, 2014) 100

4. A summary of the research instruments I used to collect data....112

5. The different sources of data I used to recruit, select, and sort participants and address each research question....113

6. An example of technology scores recorded from the classroom observations....122

7. An example of re-randomizing the treatment conditions assigned to two participants within a block....131

8. Organizing a participant's responses from the initial and exit surveys to make comparison....137

9. A breakdown of the five participants (out of the 11 who received a statistics unit) who implemented the four statistics lessons. The shaded rows correspond to the final five participants 143

10. The approximate amount of time (in minutes) each participant spent implementing Lessons 2 though 5....164

11. Karen and Robin's technology scores for Lesson 3....188

12. Karen and Robin's technology scores for Lesson 4....214

LIST OF TABLES – CONTINUED

Table Page

13. Summary of how Karen and Robin adhered to and departed from the annotations and lesson plan during their implementations of Lesson 3 232

14. Summary of how Karen and Robin adhered to and departed from the annotations and lesson plan during their implementations of Lesson 4 233

LIST OF FIGURES

Figure Page

1. The *Technological Pedagogical Content Knowledge* (TPACK) framework (http://tpack.org) 10

2. One possible relationship between statistical literacy, reasoning, and thinking (delMas, 2002). The three skills are independent of one another with some overlap 42

3. Another possible relationship between statistical literacy, reasoning, and thinking (delMas, 2002). Here, statistical literacy is viewed as the overarching goal of statistics instruction 44

4. AMTE's visual of Niess et al.'s (2005) mathematics teachers' TPACK developmental model 62

5. Lee and Hollebrand's (2011) interpretation of a *Technological Pedagogical Statistical Knowledge Framework* 63

6. Lesser and Groth's (2008) interpretation of a *Technological Pedagogical Statistical Knowledge Framework* 64

7. The experimental design I developed for this study 79

8. The traditional and integrated course pathways suggested by the *CCSSM* (NGA, 2010b) 81

9. A diagram of a completely randomized design in which participants would be randomly assigned to receive the annotated or non-annotated statistics unit 89

10. Sorting participants into four blocks based on a teacher's awareness of the use of technology and awareness of the use of inquiry in a mathematics classroom 91

LIST OF FIGURES – CONTINUED

Figure Page

11. Blocking by technology, inquiry, *and* statistical content knowledge. I created arbitrary cutoffs for the statistical content knowledge scores 92

12. Blocking by technology and inquiry, and pairing participants (within each block) based on their statistical content knowledge scores 93

13. Using the technology and inquiry rankings, along with the raw statistical content knowledge scores to sort participants into as similar of pairs as possible. Here, the size of an individual's circle corresponds to his or hers statistical content knowledge score 95

14. A continuum of the purpose of using technology and its users within the TPACK rubric 98

15. The order in which the statistics lessons were implemented and the content of each lesson....102

16. A timeline of when I collected data 111

17. Sorting participants into four blocks by their awareness of the use of technology and their awareness of the use of inquiry. The participant's statistical content knowledge scores overlay the boxplots and are represented by a unique character identifier 115

18. Statistical content knowledge scores, technology rankings, and inquiry rankings for the 28 participants initially interested in the study. Participants were ranked 1 (high) to 28 (low) on their awareness of the use of technology and their awareness of the use of inquiry. The size of the bubble surrounding a character identifier is proportional to the participant's statistical content knowledge score....118

LIST OF FIGURES – CONTINUED

Figure Page

19. The 28 participants sorted into pairs with similar rankings for technology and inquiry, as well as similar statistical content knowledge scores 119

20. A different way to create the bubble plot. Here, statistical content knowledge is the x-axis, technology ranking is the y-axis, and the size of the bubble corresponds to the inquiry ranking; smaller bubbles correspond to higher inquiry rankings 119

21. Ranking and sorting the final 11 participants into similar blocks based on technology, inquiry, and statistical content knowledge 121

22. An example of a profile plot (using hypothetical data) with *parallel* lines 126

23. An example of a profile plot (using hypothetical data) with *non-parallel* lines 126

24. An example of two participants' (hypothetical) technology scores. Their total technology score is the same, yet they display different scores in each component 129

25. Taking the difference in technology scores within each block (pair) and then using the average of those differences. An "X" represents a participant who received the annotated statistics unit and a "star" represents someone who received the non-annotated statistics unit 132

26. The attrition rate for participants in the research study 142

27. The technology rankings, inquiry rankings, and statistical content knowledge scores of the final five participants (red circles) 144

LIST OF FIGURES – CONTINUED

Figure Page

28. All five participants' *component* technology scores (offset vertically) for each lesson, separated by those who received annotations and those who did not. Lines connect the points belonging to the same participant, and participants sorted into the same block (pair) share the same color 147

29. All five participants' *total* technology score for each lesson, separated by those who received annotations and those who did not. Lines connect the points belonging to the same participant, and participants sorted in to the same block (pair) share the same color 149

30. Focusing on participants' total technology scores for Lessons 3 and 4 150

31. The change in total technology scores from Lesson 2 to 5 154

32. A figure many participants asked their students to fill in with percentages using knowledge about the empirical rule 157

33. Focusing on the one complete pair of participants (teal) who remained in the study 158

34. Participants' total technology scores for each lesson. The size of the points corresponds to the participant's inquiry ranking. Larger points indicate higher inquiry rankings 160

35. Participants' total technology scores for each lesson. The size of the points corresponds to the participant's technology ranking. Larger points indicate higher technology rankings 162

LIST OF FIGURES – CONTINUED

Figure Page

36. Participants' total technology scores for each lesson. The size of the points corresponds to the participant's statistical content knowledge scores. Larger points indicate higher statistical content knowledge scores 163

37. Participants' average time spent implementing Lessons 2 through 5 versus their average total technology score. Those who received the annotated statistics unit are represented by circles and those who received the non-annotated statistics unit are represented by triangles. Larger plotting characters correspond to higher inquiry rankings.....165

38. Comparing Karen and Robin's total technology scores for Lessons 2 through 5.....168

39. Comparing Karen and Robin's technology ranking, inquiry ranking, and statistical content knowledge scores.....171

40. A systematic approach to qualitatively analyze the field notes from classroom observations.....190

41. Karen's misconception about what the web applet was doing "behind the scenes" in Lesson 3 193

42. Karen's misconception about what the web applet was doing "behind the scenes" in Lesson 4 217

43. One resampled dataset displaying the proportion of interest.....218

44. Robin using the Smart Board to display her students' re-sampled proportions after shuffling cards by hand 221

ABSTRACT

Throughout the last twenty years, data analysis and statistics content, together with the integration of technology in mathematics classrooms, have gained increasing attention in the United States at the K-12 level. National and state standards now emphasize statistics concepts throughout high school and there is a growing motivation to shift from a traditional formula-based style of teaching statistics to a more data-oriented approach emphasizing conceptual understanding and statistical literacy. To implement this approach in the classroom, it is necessary to integrate technology into the teaching of statistics. However, many in-service high school mathematics teachers are not familiar with this process, and statistics is still a relatively new subject for most. This discrepancy highlights the need to help foster and develop in-service high school mathematics teachers' ability to effectively use technology when teaching statistics.

The goal of this study was to investigate how specially annotated lesson plans influence and guide in-service high school mathematics teachers' use of technology when teaching statistical concepts. I developed a completely randomized block experiment, using quantitative and qualitative measurements and methods of analysis. High school mathematics teachers were randomly assigned to receive an annotated or non-annotated statistics unit that included technology-based activities; four lessons were observed. The results of this study demonstrated how the process of helping teachers effectively use technology in the instruction of statistics is not straightforward; there was a large amount of variation in how teachers integrated technology and no consistent differences between the annotated and non-annotated group in this regard. All teachers, regardless of received unit, integrated technology more effectively when they were provided with a technology-based activity employing simulation. Teachers' integration of technology was most influenced by their awareness of the use of inquiry.

CHAPTER ONE

THE PROBLEM

Introduction

The teaching and learning of statistics in schools and universities has gained increasing attention over the last twenty-five years. In a world surrounded by data, the use of statistics has become ubiquitous, as results from studies using statistics have been prominently featured in the media on a daily basis. It is important, therefore, for students to learn to properly evaluate and interpret these statistical results (Franklin et al., 2007). National and state standards now strongly emphasize statistics and probability concepts in grades 6-12, and statistics has come to be seen as the new "it" Common Core subject (Bargagliotti, 2014). This has corresponded to a push from multiple organizations – such as the National Council of Teachers of Mathematics (NCTM), American Statistical Association (ASA), National Governors Association (NGA), International Commission on Mathematical Instruction (ICMI), and the International Association for Statistical Education (IASE) – to promote statistical literacy, reasoning, and thinking among all K-12 students. To help ensure that all graduating high-school students are statistically literate, these organizations recommend a "data-analytic and randomization approach" (Franklin et al., 2015) for the teaching and learning of statistics; this approach focuses on the use of real data, conceptual understanding, and a simulation-based introduction to inference, diverging from the procedure and computation of a traditional approach. For

the remainder of this document, I will refer to this "data-analytic and randomization approach" as a data-oriented approach.

To implement a data-oriented approach *and* teach statistics at a conceptual level, teachers at all levels need to have a developed content base in statistics and the pedagogical tools necessary to expose their students to experiences that develop and refine students' statistical knowledge. Central to this idea, research has shown one of the major directions for implementing a data-oriented approach for the teaching and learning of statistics involves effectively integrating technology and using it to support inquiry-based learning, engage students, and promote conceptual understanding of statistical concepts (Franklin et al., 2007, 2015).

However, most in-service K-12 mathematics teachers are not adequately prepared to implement this new approach (Lazar & Franklin, 2015). The subject of statistics is relatively new to many teachers (Franklin et al., 2007), and they have indicated a lack of confidence and training in teaching it, working with data, and effectively integrating technology into their mathematics classrooms (Franklin, 2013; Lazar & Franklin, 2015; U.S. Department of Education, 2016). Teachers' readiness and teaching philosophies affect how well technology is implemented in their classrooms (Niess, Browning, Driskell, Johnston, & Harrington, 2009). Teachers who did not learn with technology may be predisposed towards not teaching with technology (Batanero, Burrill, & Reading, 2011). Teachers have also stated they lack the time needed to invest in learning how to effectively use technology and develop technology-based activities (Wachira & Keengwe, 2011). The challenge, therefore, is to identify methods of technology-assisted

instruction that foster and develop in-service K-12 mathematics teachers' proper and effective use of technology in the context of teaching statistical concepts.

As Hiebert and Morris (2012) have argued, directly focusing on the methods used to teach rather than the individual characteristics of teachers themselves is an alternative means to improve teaching. Relevant knowledge for teaching may be externally stored and passed along in artifacts – instructional materials (such as specially annotated lesson plans) that are both readily available to all instructors and able to be improved as researchers and educators implement and revise them (Morris & Hiebert, 2011). Therefore, developing, implementing, refining, and sharing such annotated lesson plans may be an appropriate means to guide teaching statistics with technology in ways that continually foster and develop all teachers' abilities to effectively use technology to teach statistical concepts.

Background

As available technology and the world around us have evolved, so too have statistics curricula at the K-12 level. With the development and adoption of the *Common Core State Standards for Mathematics* (CCSSM) and other state standards, the subject of statistics now comprises a substantial area of study within a K-12 mathematics curriculum (Peck, Gould, Miller, Wilson, & Zbiek, 2013). Data and information today are broader and more accessible than ever, and thus statistical literacy is essential for citizens to be informed and productive members of society. The applicability of statistics to problems people face on a daily basis, coupled with the use of statistical inference in

promoting individuals' capacity for critical thinking and reasoning, are two of the key reasons for present-day education to focus on engaging students in fully understanding statistical inference and its associated processes (Franklin et al., 2015).

The emergence of statistics (often labeled "Data Analysis and Probability") as a vital subject of study began with NCTM's *Curriculum and Evaluation Standards for School Mathematics* (1989) and persisted as these standards were updated in 2000 with the *Principles and Standards for School Mathematics*; in 2007, with the release of ASA's *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report*; in 2010 with the *CCSSM*; and, most recently with the release of ASA's *Statistical Education of Teachers* (SET) *Report*. Together, these documents have been instrumental in further developing statistics education in the K-12 mathematics curricula. They present a data-oriented approach for teaching and learning statistics to promote conceptual understanding and statistical literacy, reasoning, and thinking; this stands in stark contrast to the rote memorization within traditional instruction (Ben-Zvi & Garfield, 2004). Importantly, the use of a data-oriented approach engages students in the four primary processes of statistical investigation: formulating questions, collecting data, analyzing data, and interpreting results.

The *CCSSM*, *Pre-K-12 GAISE Report* and *SET Report* emphasize the idea of statistical inference – a concept useful in understanding what one can conclude about the world using a wealth of surrounding data. Inference, in short, is making a broader conclusion from available data or premises known or assumed to be true. For example, a fisherman catches ten albino trout in one (productive) evening. Using the information

from the ten, she concludes the majority of fish in the lake are albino trout, though there is clearly uncertainty in her conclusion based on the ten fish observed that evening. *Statistical* inference formally links the data to one's uncertainty in conclusions through a probability model. Extending this idea to the previous example, a researcher *randomly* selects fish from different parts of the lake; they are all albino trout. The researcher concludes (with a quantified amount of uncertainty) the majority of fish in the lake are albino trout.

The provided examples illuminate Ramsey and Schafer's (2013) definitions of inference and statistical inference. "An inference is a conclusion that patterns in the data are present in some broader context. A statistical inference is an inference justified by a probability model linking the data to the broader context" (p. 8). Students often have difficulty with the concept of statistical inference; for example, they struggle understanding why one can draw conclusions from a sample representing a fraction of a larger population (Peck et al., 2013). Fundamental to the understanding of the machinery underlying statistical inference are statistical literacy, reasoning, and thinking.

While multiple organizations promote statistical literacy, reasoning, and thinking, there are no consistent definitions for these three components. Often, the three are all used interchangeably (Ben-Zvi & Garfield, 2004; Chance, 2002), though many researchers have made an effort to define and distinguish between them. The first *College GAISE Report* (Garfield et al., 2005) defined statistical literacy as "understanding the basic language of statistics (e.g., knowing what statistical terms and symbols mean and being able to read statistical graphs) and fundamental ideas of statistics" (pg. 14). In

general, a statistically literate high-school graduate will be able to evaluate and understand conclusions presented in a scientific study and make informed decisions about the study (Franklin et al., 2007). One may think of statistical reasoning as "the way people reason with statistical ideas and make sense of statistical information" (Ben-Zvi & Garfield, 2004). Here, statistical reasoning refers to the understanding and ability to explain statistical processes and interpret results. Statistical thinking, on the other hand, involves an "understanding of why and how statistical investigations are conducted" (Ben-Zvi & Garfield, 2004). In particular, one must understand the concepts underlying statistical investigations – such as the nature of variability and sampling, and why random assignment in an experiment can establish causation (Ben-Zvi & Garfield, 2004). This focus on variability sets mathematics apart from statistics, a distinction the *Pre-K-12 GAISE Report* (Franklin et al., 2007) recommends teachers be aware of.

Although statistics is a content domain that is widely accepted, one does not typically consider statistics a separate discipline from mathematics at the K-12 school-level despite several notable qualities that distinguish the two (Franklin et al., 2007, 2015; delMas, 2004). Preparing teachers to teach statistics requires a different approach from preparing teachers to teach mathematics (Garfield & Ben-Zvi, 2008). For this end, mathematics and statistics researchers and educators have collaborated to ascertain how to train the teachers ultimately responsible for teaching statistics at this level. Appropriately, the *SET Report* (Franklin et al., 2015), the *Pre-K-12 GAISE Report* (Franklin et al., 2007), the *Mathematical Education of Teachers* (MET) *Reports I* and *II* (CBMS, 2001, 2012), and the *Proceedings of the Joint ICMI/IASE Study Conference*

(Batanero, Burrill, Reading, & Rossman, 2008) have all helped raise awareness of the challenges faced in training and supporting K-12 mathematics teachers who teach statistics. To teach statistics effectively, teachers must have a deeper understanding of the statistical concepts they expect their students to know and learn (Peck et al., 2013). Teachers ought to understand different representations of given problems and be able to anticipate and address the misconceptions and difficulties students face (Peck et al., 2013); they must also use appropriate technological tools to teach statistics in ways that support inquiry-based learning, engage students and enhance their conceptual understanding of statistical concepts.

Research studies focusing on the statistical preparation of teachers lend insight into the current state of teachers' readiness to teach statistics. Many in-service K-12 mathematics teachers have not been trained to teach statistics at the level suggested by the *Pre-K-12 GAISE Report* and the *CCSSM* (Batanero et al., 2011; Bargagliotti, 2014; Franklin, 2013). Most mathematics teachers are not comfortable with the subject of statistics and lack confidence in their instruction, often due to a relative lack of statistics education (Peck, Kader, & Franklin, 2008; Franklin & Kader, 2010); even some of those who majored in mathematics have never taken a data-oriented introductory statistics course (Batanero et al., 2008, 2011; Franklin, 2013). Additionally, these reports have identified limitations in present-day instruction of statistics. It is natural to think that a mastery of statistics comes primarily from knowing the processes used in collecting and analyzing data. However, teachers need a deeper understanding than this to teach

statistics; they must also be aware of the complex interactions among these processes so they can be well-communicated to students (Peck et al., 2013).

Students being taught under a traditional, formula-based approach are not gaining a rich understanding of the statistical concepts and are not equipped to reason and think at a deep level statistically (Ben-Zvi & Garfield, 2004). Accordingly, successful implementation of the national and state statistics content standards requires teachers to deviate from the comfort of traditional approaches and focus more on conceptual understanding and statistical literacy, thinking, and reasoning. These statistics content standards are intended to introduce statistical concepts on a more intuitive level, using actual data married with computer simulations (NGA, 2010a).

The recommended data-oriented approach to teach and learn statistics would not be possible without the use of technology. Technological innovations enable teachers to instruct students in new and effective ways that were not feasible (or possible) in the past. Technology has allowed teachers to work with, visualize, and explore real datasets as well as rely on computer simulations to teach and learn many statistical concepts (Chance, Ben-Zvi, Garfield, & Medina, 2007). However, the newfound prevalence of technology presents another challenge for teaching statistics. In addition to learning the content they will ultimately teach, most teachers now need training in teaching statistics with technology. They must learn the skills and practices necessary to teach with technology at the levels suggested by the *CCSSM*, *Pre-K-12 GAISE Report*, and *SET Report*.

The effectiveness of a teacher's use of technology to teach statistics is based on factors including a teacher's statistical content knowledge, his or her base of knowledge in the use of technology, particularly in supporting statistics instruction, and an understanding of pedagogy as it relates to statistics education (Koehler & Mishra, 2005; Lee & Hollebrands, 2011). In order for these teachers to implement instructional techniques that maximize the scope of knowledge a student can be exposed to and understand fully, it is simply not enough for teachers to learn how to use current types of technology. Teachers must also develop new techniques and skills to teach with technological tools in ways that help students develop conceptual understanding of the material.

To illustrate this idea, Koehler and Mishra (2005) introduced the term *technological pedagogical content knowledge* (TPACK), describing the range of teachers' knowledge in the process of effectively integrating technology into their teaching, with the understanding that pedagogy, content, and technology interrelate in a way that is fundamental to the process of developing effective ways of teaching with technology (Figure 1). The TPACK framework offers a baseline for teachers' training in technology and is guided by the observation that knowledge of technology alone is not enough to effectively integrate technology in the classroom. Instead, teachers ought to use all three components – technological knowledge, pedagogical knowledge, and content knowledge – to effectively integrate technology into their instruction.

Teachers should not just use technology solely for the sake of using it. They must be able to identify the value of a technological tool that uniquely enables its application

to individual situations in learning. Teachers, thus, should know how to use a given technological tool, though they must also criticize with its use to fully determine its role in inquiry-based instruction.

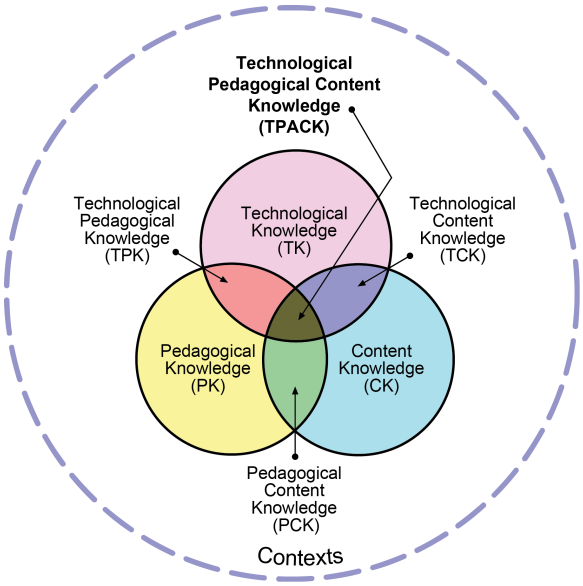


Figure 1. The *Technological Pedagogical Content Knowledge* (TPACK) framework (http://tpack.org).

It is a widely-held assumption in the U.S., documented by Hiebert and Morris (2012), that one must focus primarily on improving teachers to improve teaching. Acting on this assumption, educators and policymakers have attempted to recruit more talented people with the "right" characteristics and engrain desirable traits in others through pre-service teacher education programs and professional development workshops. Hiebert and Morris (2012) propose an alternative approach to improve teaching. Rather than focusing on the qualities of the people who teach, they suggest focusing "directly on the methods used to teach" (Hiebert and Morris, 2012).

One form this approach can take is in the use of specially annotated lesson plans. Hiebert and Morris (2012) claim annotated lesson plans can act as a store of knowledge gathered for more effective methods of teaching. These can be effectively implemented by different teachers in different classrooms and allow for the transmission of knowledge between generations of teachers to support continuous improvement. There is a great need to develop these high-quality instructional materials for teachers, particularly to support the development of teachers' TPACK in statistics (Lee & Hollebrands, 2008).

Statement of the Problem

Teaching and learning statistics should model the practice of statistics, in which technology is an indispensable tool. The NCTM places a special emphasis on the role of technology in the classroom. They argue "technology is essential in teaching and learning mathematics; it influences the mathematics that is taught and enhances students' learning" (NCTM, 2000). Furthermore, the *Pre-K-12 GAISE Report* (Franklin et al., 2007) states

> advances in technology and modern methods of data analysis in the 1980s, coupled with the data richness of society in the information age, led to the development of curriculum materials geared toward introducing statistical concepts into the school curriculum as early as the elementary grades (pg. 3).

Meanwhile, the most recent *College GAISE Report* (Carver et al., 2016) recommends the use of technology to explore concepts and analyze data, arguing that innovation in technology makes statistical procedures more accessible. Hence, time regularly spent performing calculations can be freed to allow for "greater emphasis on understanding concepts and interpreting results" (pg. 16).

Among the 2016 *College GAISE Report's* six recommendations for teaching statistics is the suggestion that the proper and effective use of technology aids in developing conceptual understanding and analyzing data. "Technology should aid students in learning to think statistically and to discover concepts. It should also facilitate access to real (and often large) datasets, foster active learning, and embed assessment into course activities" (pg. 19; Carver et al., 2016). The report also suggests teachers perform simulations to illustrate abstract concepts. Likewise, the *CCSSM* address the use of technology and simulation; for example, the statistics content standard HSS-IC.B.5 requires students to "use data from a randomized experiment to compare two treatments" and "use simulations to decide if differences between parameters are significant" (NGA, 2010a).

The issue of the proper and effective role of technology in the instruction of statistics was raised during the 2011 Joint ICMI/IASE Study Conference, where it was recognized as one of the key challenges to address in the statistics curricula. Recently, the *Pre-K-12 Report, College GAISE Report*, and *SET Report* have addressed this challenge by providing recommendations and suggestions for teaching statistics with technology. The *SET Report* supports the recommendation, lain out in the *Pre-K-12* and *College GAISE Reports*, that teachers should use technology mindfully for specific ends (like developing concepts and analyzing data). For example, the *College GAISE Report* (Carver et al., 2016) suggests teachers employ Shiny apps coded in **R** to explore statistical concepts, and the *SET Report* (Franklin et al., 2015) recommends technologies

that “have the capability to create dynamic graphical displays, produce numerical summaries of data, and perform simulations easily” (pg. 24).

As statistics is not taught independently, but rather as part of a K-12 mathematics curriculum, it is essential for K-12 mathematics teachers to be prepared to teach it with technology. Generally, depending on a given teacher’s TPACK, technology may either enhance or hinder students’ learning and conceptual understanding. TPACK, therefore, is necessary to effectively use technology to teach statistics at the K-12 level and prepare those students for the level of rigor required in a college statistics course or the challenges presented in entering the workforce.

Research Questions

The purpose of this study was to examine the degree to which specially annotated lesson plans influence in-service high school mathematics teachers’ integration of technology when teaching statistics. I investigated three specific research questions, each representing a part of this broad purpose.

1. To what extent do specially annotated lesson plans enable in-service high school teachers to effectively integrate technology in their instruction of statistical concepts?
2. Do the qualities inherent in specially annotated lesson plans carry over into in-service high school teachers’ implementation of a future technology-assisted statistics lesson?

3. How are the following qualities associated with an in-service high school teacher's use of specially annotated lesson plans as a means to guide the instruction of statistics with technology?
    a) The teacher's statistical content knowledge;
    b) the teacher's awareness of the use of inquiry in a mathematics classroom;
    c) the teacher's awareness of the use of technology in a mathematics classroom; and
    d) how the teacher's implementation of a specially annotated lesson plan aligned with its goals.

Importance of the Study

The primary goal of the *Pre-K-12 GAISE Report* (Franklin et al., 2007) is to provide guidelines to achieve the statistical literacy of all students; the document asserts "every high-school graduate should be able to use sound statistical reasoning to intelligently cope with the requirements of citizenship, employment, and family and to be prepared for a healthy, happy, and productive life" (pg. 1). The *CCSSM's* fifth *Standard for Mathematical Practice* requires students to "use appropriate tools strategically," with proficient students being able to "use technological tools to explore and deepen their understanding of concepts" (pg. 7; NGA, 2010a). For students to be able to use technology appropriately when learning statistics, they must be taught how to do so. As such, teachers need to have a deep understanding of statistics, an ability to effectively use

appropriate technological tools in the classroom to support inquiry-based learning, promote statistical literacy and conceptual understanding, and an ability to connect these in a meaningful way.

In a world and discipline emphasizing the use of technology, teachers cannot truly be successful instructing their classes in statistics without the knowledge and expertise needed to effectively integrate technology into their classrooms; teachers with this ability offer their students a considerable advantage in learning statistics. However, in-service K-12 mathematics teachers typically have not been exposed to experiences that develop their ability to effectively integrate technology. Even veteran teachers are often limited in their breadth of TPACK-developing experiences. Teachers – both novice and experienced – should be presented with particular experiences that foster their ability for using technology in their respective classrooms (Niess, Suharwoto, Lee, & Sadri, 2006). Accordingly, researchers and teacher educators providing professional development to in-service teachers are interested in understanding what kinds of support teachers need to develop and strengthen their ability to effectively integrate technology into their classes.

This study will inform educators and researchers about the influence annotated lesson plans may have on in-service teachers' use of technology for statistics instruction. With this knowledge, researchers and teacher educators can make decisions about the type of professional development to be offered to in-service teachers to improve teaching statistics with technology. Teacher educators may also use annotated lesson plans as a means to improve the preparation of pre-service teachers. Ultimately, those K-12 students

who have teachers that can effectively teach statistics with technology will benefit in their development of statistical literacy, reasoning, and thinking skills.

List of Definitions and Acronyms

The following list of definitions and acronyms used throughout this dissertation are defined below.

AMTE	Association of Mathematics Teacher Educators
ASA	American Statistical Association
CCSSM	Common Core State Standards for Mathematics
Data-oriented approach	"A modern data-analytic and randomization approach to teaching statistics. This involves the use of a simulation-based introduction to inference and focuses on conceptual understanding rather than computation and procedural understanding" (Franklin et al., 2015).
GAISE	Guidelines for Assessment and Instruction in Statistics Education
IASE	International Association for Statistical Education
ICMI	International Commission on Mathematics Instruction
In-service teacher	A teacher who is currently employed at the Pre-K-12 level.
Inference	"A conclusion that patterns in the data are present in some broader context" (Ramsey & Schafer, 2013).
MET	Mathematical Education of Teachers
NCTM	National Council of Teachers of Mathematics
NGA	National Governors Association
Pre-service teacher	A college student who is training to be a Pre-K-12 teacher.

SET	Statistical Education of Teachers
Simulation	A method, using an assumed probabilistic model, to generate random outcomes that resemble real world outcomes.
Statistical inference	"An inference justified by a probability model linking the data to the broader context" (Ramsey & Schafer, 2013).
Student	A child or adolescent in a Pre-K-12 classroom.
Teacher	A current or prospective teacher in a Pre-K-12 classroom.
Teacher educator	Educators who further the education of pre- and in-service teachers through instruction, professional development, or research.
TPACK	Technological Pedagogical Content Knowledge

CHAPTER 2

REVIEW OF THE LITERATURE

Introduction

The primary goal of this chapter is to review the literature – namely, of statistics education, the integration of technology into the teaching and learning of mathematics and statistics, and the use of annotated lesson plans to improve teaching. In the first part of this chapter, I present the history and current state of K-12 statistics education with special consideration of the influence of national documents. As the goals of these documents are aligned with the goal of producing statistically literate members of society who can think and reason about data, I offer a brief history of statistical literacy, reasoning, and thinking. Next, I examine the relationship between statistics and mathematics, elaborating on reasons statistics is taught collectively rather than independently under the prevailing K-12 mathematics curriculum. Following this, I discuss the literature regarding the teaching and learning of statistics, particularly with technology and focusing on the challenges faced by K-12 mathematics teachers. Next, I review the description of the TPACK framework and development model, identifying different metrics used to measure a teacher's TPACK. I then discuss the potential of annotated lesson plans as a means to improve teaching. Finally, I conclude this chapter by establishing the theoretical framework used for this study.

History of K-12 Statistics Education

Although the recent prominence of statistics at the K-12 level is, in part, the direct result of the 1980s reform movement of pre-college and college statistics education, the suggestion that statistics be included in the school curriculum is not a new idea (Scheaffer & Jacobbe, 2014). Prompted by the demands of industry and business, knowledge of basic statistics became a mathematics requirement at the K-12 level. In 1923, the Mathematical Association of America released the *Reorganization of Mathematics in Secondary Education Report*, recommending the inclusion of statistics at the junior high and high school level (National Committee on Mathematical Requirements, 1923). In the decade following World War II, statistics again came to be seen as an important subject to teach, as it was used heavily during the war and periods of expansion thereafter. The 1960s and 1970s did not see much progression towards including statistics as part of the school curriculum, though the emphasis of statistics education began to change from statistical theory to the practical analysis of data (Scheaffer & Jacobbe, 2014). Prior to this, the focus of statistics instruction was in rote memorization, formulas, and calculations. By the 1980s, following a shift in the statistics community from theory to data analysis, the necessity of including statistics in education for both pre-college and college students became clear (Garfield & Ben-Zvi, 2008; Scheaffer & Jacobbe, 2014). This perspective was subtly redefined in the 1990s and continues into the present day, as traditional instruction became increasingly regarded as ill-equipped to produce statistically literate citizens (Franklin et al., 2015; Scheaffer & Jacobbe, 2014; Franklin, 2013).

The reform movement starting in the 1980s established the importance of statistical thinking and statistical processes – particularly, in analyzing data and defending one's results and conclusions. The advances in technology and new focus on developing statistical literacy has given way to K-12 curriculum materials and reports; these have been geared toward shaping conversations about mathematics and statistics education, working toward ensuring every high school graduate is well-prepared for college or the workforce.

The first of these reports was the NCTM's *Curriculum and Evaluation Standards for School Mathematics*, released in the late 1980s. This document preceded the NCTM's 2000 revision of the standards (the *Principles and Standards for School Mathematics*). Next, the ASA published the *Pre-K-12 GAISE Report* in 2007; the release of the *CCSSM* followed this in 2010, and recommendations for preparing teachers to teach mathematics (*MET II Report*) and statistics (*SET Report*) were released in 2012 and 2015, respectively.

Curriculum and Evaluation Standards for School Mathematics

During the late 1980s, the NCTM developed and released the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989), commonly known as the *Standards*, to help improve the quality of K-12 mathematics and shift the focus from memorization and computation to an inquiry-oriented classroom environment emphasizing conceptual understanding and problem solving. The content standards are descriptions of what students should know and be able to do, spanning three grade-level

groups (K-4, 5-8, and 9-12). Statistics has an integral part in each grade band, as it is meant to be integrated with other mathematics topics. Under the statistics standards for grades 9-12, the focus lies on collecting, representing, summarizing, and analyzing data while applying this knowledge to evaluate statistical claims. Eight goals for students were delineated, focusing primarily on the understanding of concepts, data analysis, and communication. Specifically, the *Standards* state all students should

- construct and draw inferences from charts, tables, and graphs that summarize data from real-world situations;
- use curve fitting to predict from data;
- understand and apply measures of central tendency, variability, and correlation;
- understand sampling and recognize its role in statistical claims;
- design a statistical experiment to study a problem, conduct the experiment, and interpret and communicate the outcomes; and
- analyze the effects of data transformations on measures of central tendency and variability (NCTM, 1989).

From the start, the NCTM has seen its involvement as a continual process of improvement. It has noted that for its standards to continue to be relevant, they must be criticized, tested, and changed, if necessary (NCTM, 2000).

Principles and Standards for School Mathematics

In 2000, the NCTM published the *Principles and Standards for School Mathematics* (PSSM; NCTM, 2000), a revision to its 1989 *Standards*. Similar to the 1989 *Standards*, the *PSSM* present a common foundation of mathematics to be learned by all pre-K-12 students, building upon the messages of the previous document (NCTM, 2000). This document contains *Principles* (themes that are important to mathematics education) and *Standards* (descriptions of what students should know and be able to do). The two combined "constitute a vision to guide educators as they strive for the continual improvement of mathematics education in classrooms, schools, and educational systems" (NCTM, 2000). Part of the revisions include an expansion of scope: pre-kindergarten grades were a new addition. The *PSSM* now span four grade levels (Pre-K-2, 3-5, 6-8, and 9-12) where "*Statistics and Probability*" standards are combined under the umbrella of "*Data Analysis and Probability*." This is an improvement on the old 1989 *Standards*, in that the *PSSM* more clearly describe the expectations of students to formulate research questions and collect, summarize, analyze, and present the necessary data needed to investigate them. For all students (not just those in grades 9-12), the stated statistics and probability expectations are to

- formulate questions that can be addressed with data, and collect, organize, and display relevant data to answer them;
- select and use appropriate statistical methods to analyze data;
- develop and evaluate inferences and predictions that are based on data; and
- understand and apply basic concepts of probability (NCTM, 2000).

These major areas of expectation were detailed further into 19 particular standards, enabling teachers and students to have a better understanding of the content both need to know and the tasks both must be able to do.

The *PSSM* catalyzed one of the first formal movements in statistics education, again reemphasizing the key role of statistics in K-12 mathematics education (Peck et al., 2008; Franklin, 2013). This document prompted the development of future documents about K-12 statistics education (Rossman & Franklin, 2013; Franklin et al., 2007; Peck et al., 2008), and particularly the *Pre-K-12 GAISE Report*.

Guidelines for Assessment and Instruction in Statistics Education

To address the growing importance of statistics at the pre-K-16 level and the need to provide guidance to teachers implementing statistics, the ASA released two reports offering a set of proposed guidelines for instruction and assessment in pre-college and college education. The first, the *College GAISE Report* (Garfield et al., 2005) focuses on the goals for students in a college introductory statistics course. With recent advances in statistics education and technology, this report was recently revised and released in July 2016 (Carver et al., 2016). The second, and arguably most influential document of the early 21st century in K-12 statistics education (Scheaffer & Jacobbe, 2014), is the *Pre-K-12 GAISE Report* (Franklin et al., 2007).

The guidelines in the *Pre-K-12 GAISE Report* were written with implementing the "*Data Analysis and Probability*" content standards (as in the *PSSM*) in mind (Groth & Bargagliotti, 2012). As a document, it accompanies the *PSSM's* recommendations,

allowing for greater ease of interpretation of the standards (Rossman & Franklin, 2013; Franklin et al., 2007). The *Pre-K-12 GAISE Report* presents the statistics curriculum for grades Pre-K-12 as a "conceptual, cohesive, and coherent curriculum strand" (Franklin & Kader, 2010; Groth & Bargagliotti, 2012). Pedagogical approaches for teaching statistics and likely progressions for students' education in statistics are also included.

The main goal of the *Pre-K-12 GAISE Report* is to provide guidelines to help students be statistically literate by the end of their pre-college education (Franklin et al., 2015; Bargagliotti & Franklin, 2015). To accomplish this, the report emphasizes variability, delineates the differences between mathematics and statistics, and describes a useful two-dimensional framework for learning statistical concepts (Franklin et al., 2007). One dimension of the framework is defined by the "problem-solving process components" plus the nature and role of variability; the second contains three levels of development (Franklin et al., 2007).

Particularly notable in the *Pre-K-12 GAISE Report* is the decreased focus on statistical theory and formulaic exercises (Scheaffer & Jacobbe, 2014). The guidelines in this report are developed around data analysis; data analysis is, of course, central to the idea that statistical problem solving is an investigative four-step process of formulating a statistical question, collecting and analyzing appropriate data, and then drawing conclusions (Franklin et al., 2007). This is consistent with the four stated statistics and probability expectations prescribed by the *PSSM*.

The second dimension of the *Pre-K-12 GAISE Framework* describes how each step of this investigative process encompasses three developmental levels, generally

corresponding to elementary, middle, and high school grade levels, respectively (Franklin et al., 2007). Greater difficulty and complexity occurs with progression through the developmental levels (Franklin et al., 2007; Bargagliotti & Franklin, 2015). Unlike the NCTM's standards (NCTM, 1989, 2000), students progress through the levels only as a result of greater sophistication in statistical thinking – *not* age (Franklin et al., 2007; Peck et al., 2008).

The *Pre-K-12 GAISE Report* provides information about implementing the statistics content standards as stated in the *PSSM* and *CCSSM,* and it provides examples showcasing the complexity of statistical thinking over the developmental levels (Franklin et al., 2007). Thus, educators can use this report as a guide to implement the statistics content standards, most notably from the *CCSSM*. It affords itself great utility as a companion to the *CCSSM*, as it describes many ideas not contained in the *CCSSM*. For instance, it contains pedagogical approaches for statistics, meaningful statistical connections, and developmental trajectories for students' statistical thinking (Groth & Bargagliotti, 2012). Ultimately, the *Pre-K-12 GAISE Report* influenced the writers of the *CCSSM* greatly (Scheaffer & Jacobbe, 2014; Franklin et al., 2015).

Common Core State Standards for Mathematics

In 2009, the Common Core State Standards were developed in an effort organized at the state level. The primary purpose of this movement was to create a basis of standards, shareable nationwide, that would enable a level of rigor and completeness not previously seen in the mathematics curricula (NGA, 2010a; Bargagliotti, Jacobbe, &

Webb, 2014). Eventually released in 2010 by the NGA Center and the Council of Chief State School Officers, 42 states (as of November 2016) have adopted and started implementing the *CCSSM*. The *CCSSM* are formed from two key components: standards for mathematical practices and standards for mathematical content. The former are eight mathematical practices and mathematical habits of mind, while the latter describe mathematics and statistics concepts that should be taught and learned at the K-12 level.

Similar to the NCTM's *Standards*, and in contrast with the *Pre-K-12 GAISE Report*, the *CCSSM* avoid prescribing pedagogy (Groth & Bargagliotti, 2012). It is a national policy document establishing the end result (what K-12 students should have learned) but not the route teachers should take to arrive there (NGA, 2010a). Although the *Standards*, *PSSM* and the *Pre-K-12 GAISE Report* emphasize statistics to be taught at all levels through grades K-12, there is little emphasis on statistics in grades K-5 in the *CCSSM*. In contrast, the *CCSSM's* "*Statistics and Probability*" content standards are first heavily introduced and developed in grades 6-8 and then further developed throughout high school (Franklin et al., 2015).

The *CCSSM* indicate expectations for K-12 students. In particular, high school students must be able to

- summarize, represent, and interpret data on a single count or measurement variable;
- summarize, represent, and interpret data on two categorical and quantitative variables;
- interpret linear models;

- understand and evaluate random processes underlying statistical experiments; and
- make inferences and justify conclusions from sample surveys, experiments, and observational studies (NGA, 2010a).

Some observers have commented on how the basis for an Advanced Placement (AP) Statistics or college introductory statistics course is shared with the statistics content standards in the *CCSSM* (Franklin et al., 2015; Franklin, 2014). However, Franklin (2014) states "the statistics standards at the K-12 level are intended to conceptually introduce basic statistics topics utilizing case studies, technology, and simulation." That is, these standards intend for basic statistical ideas to be presented more intuitively than how they are presented in an AP Statistics or college introductory statistics course (Franklin, 2014; Franklin et al., 2015).

The statistics standards in the *CCSSM* reflect the investigative statistical process described in the *Pre-K-12 GAISE Report* and revolve around developing a conceptual understanding of statistical topics (Franklin et al., 2015). As a result of these more demanding recommendations, the Conference Board of Mathematical Sciences (CBMS) acknowledged teachers' lack of preparation in statistics in the *Mathematics Education of Teachers* (MET) *II Report* (CBMS, 2012).

The Mathematical Education of Teachers

The CBMS is an organization containing seventeen professional groups; among these are the Mathematical Association of America and the American Mathematical

Association. All of the groups involved in the CBMS have the pursuit and spread of knowledge in mathematics as one of their stated purposes. The broad goal of the CBMS is to enable its component groups to work together so research may be conducted to improve the quality of education for all.

The *MET I* and *II Reports* (CBMS, 2001, 2012), written by mathematicians, statisticians, and other experts, focus on the preparation of K-12 mathematics teachers; these are resources for pre- and in-service teachers as well as teacher educators. These reports suggest the depth and scope of knowledge teachers ought to know and how they should go about learning it (CBMS, 2012). Recently, in light of the release of the *CCSSM* and *Pre-K-12 GAISE Report*, the CBMS updated its *MET I Report* (CBMS, 2001) to the *MET II Report* (CBMS, 2012). This document identifies the acute need for preparing teachers in statistics. In particularly, the *MET II Report* (CBMS, 2012) states

> in preparation for teaching this [statistics and probability], teachers should see real-world data sets, understand what makes a data set good or bad for answering the question at hand, appreciate the omnipresence of variability, and see the quantification and explanation of variability via statistical models that incorporate variability (p. 58).

It is, of course, necessary for teachers to be given a modern version of a statistics course – one that "centers around statistical concepts and real-world case studies, and makes use of technology in an active learning environment" (CBMS, 2012).

Although the *MET II Report* offers some guidelines for training K-12 teachers to teach statistics, more vital to this report is the *mathematics* education of teachers. Correspondingly, the ASA-NCTM Committee identified the need for a companion document (titled the *Statistical Education of Teachers*, or SET), building upon the *MET*

II Report and providing more detailed guidance in regards to K-12 *statistics* education (Franklin et al., 2015; Franklin, 2014; Scheaffer & Jacobbe, 2014).

The Statistical Education of Teachers

In April of 2013, the ASA and NCTM released a joint position statement addressing the question of preparation and support K-12 teachers need to effectively support students' learning of statistics:

> Students, teachers, administrators, employers, and others increasingly recognize statistics as an important scientific field of study. Mathematics content standards emphasize that the development of statistical thinking begins in the early grades and extends into and beyond high school. To successfully develop students' statistical thinking, teachers must have deep knowledge and understanding of statistics and the way that students learn statistics. Consequently, the need is critical for high-quality preservice and in-service preparation and professional development that supports pre-K-12 teachers of mathematics, new and experienced, in developing their own statistical proficiency as well as their students' understanding of and skill in working with statistics (p. 1; ASA-NCTM, 2013).

In response to the *MET II Report* and the growing need to train teachers to teach statistics, the ASA released the *SET Report* (Franklin et al., 2015). This represented a bold effort to clearly support the statistical preparation of both pre- and in-service teachers at the level envisioned by the *Pre-K-12 GAISE Report* and *CCSSM* (Franklin et al., 2015; Bargagliotti & Franklin, 2015). The *SET Report* expands upon previous K-12 statistics standards, outlining concepts in a straightforward manner at each grade level (elementary, middle, and high school) to ensure statistical literacy for all (Franklin, 2014; Lazar & Franklin, 2015); the document is clear to distinguish statistical preparation from mathematical preparation. Furthermore, it highlights the *CCSSM's* eight mathematical practices viewed through a statistical lens (Franklin et al., 2015) and provides examples

illustrating key concepts, themes, and difficulties. Here, the mathematical practices are stated in terms of what a *statistically* proficient student should be able to do.

The *SET Report* serves as a resource to help "prepare teachers to teach the intellectually demanding statistics in the PreK-12 curriculum" (Franklin et al., 2015). This report has three main goals: first, to reach the appropriate audience preparing teachers of statistics, second to offer recommendations for the breadth and depth of statistical content necessary for K-12 students to learn, and finally, to offer solutions for the difficulty of combining each teacher's statistical training with his or her broader training in education (Franklin, 2014). The *SET Report* aims to accomplish the following:

- clarify MET II's recommendations for the statistical preparation of teachers at all grade levels: elementary, middle, and high school;
- address the professional development of teachers of statistics;
- highlight differences between statistics and mathematics that have important implications for teaching and learning;
- illustrate the statistical problem-solving process across levels of development; and
- make pedagogical recommendations of particular relevance to statistics, including the use of technology and the role of assessment (Franklin et al., 2015).

As noted by the *MET* and *SET Reports*, the preparation of teachers should differ between mathematics and statistics because they require a different style of learning. Students and

teachers should respond differently to these two subjects (Rossman, Chance, & Medina, 2006).

Statistics versus Mathematics

Research has discussed how statistics and mathematics are fundamentally different disciplines (Franklin et al., 2007, 2015), although there is some overlap. Statistics uses mathematics, and probability is often a link that ties them together (Groth, 2007; delMas, 2004). Solutions to some mathematics problems can be *deduced* using probability theory, whereas a solution to a statistics problem can be *inferred* using probability theory (Franklin et al, 2007). It is the distinction between deduction and inference that fundamentally distinguishes between mathematics and statistics.

Statistics is a scientific field of study all on its own, with its own core concepts and ideas and with a foundation rooted in mathematics (Leavy, Hannigan, & Fitzmaurice, 2013). Hence, it is often considered a mathematical science. As a result, at the K-12 level, statistics has naturally been incorporated into the mathematics curriculum and taught as a subtopic of mathematics (Franklin et al., 2015). People often mistakenly expect statistics to be like mathematics, with an emphasis on computation and formulas; students and teachers often equate statistics with mathematics and approach the discipline of statistics with the same mindset, skills, and tools they use for mathematics (Garfield & Ben-Zvi, 2008). However, statistics and mathematics each requires a different style of cognition (Franklin et al., 2015; CBMS, 2012). “Mathematics is built on a foundation of deduction

proof, and truth, whereas statistics is based on inference, data analysis, and uncertainty" (Peck et al., 2013).

One of the major differences between mathematics and statistics is the role of context. Numbers in statistics (i.e., data) are not abstract and isolated; they come inextricably linked with a particular context with a particular method for collecting data (Franklin et al., 2015; Leavy et al., 2013; Scheaffer, 2006). Although many mathematicians lean on applied contexts in their own research and teaching, patterns that are not grounded in the application often remain a focus. In this sense, the context of the mathematics problem is a degree removed from the focus, serving to motivate the problem (Cobb & Moore, 1997). Mathematics, in all of its abstraction, removes context – the very thing that, in statistics, defines the scope of data so they can be analyzed in a way that is coherent and useful, ultimately driving statistical inference.

Another distinction between mathematics and statistics is the focus on variability (Franklin et al., 2015). Variability is central to statistics, and the *Pre-K-12 GAISE Report* (Franklin et al., 2007) acknowledges it at all levels of the four-step statistical investigative process. Problem solving in statistics relies on measuring and understanding the nature of variability found in the data (Kader & Perry, 2006). Mathematics, on the other hand, primarily focuses on answering deterministic questions (Franklin et al., 2015; Groth, 2007; Scheaffer, 2006).

To help distinguish the differences between mathematics and statistics, two questions are presented below (Scheaffer, 2006). These questions were listed as "*data analysis*" questions on a previous National Assessment of Educational Progress (NAEP)

mathematics assessment. They are "fairly typical of questions that find their way onto assessments in statistics" (Scheaffer, 2006).

Question 1:
The average weight of 50 prize-winning tomatoes is 2.36 pounds. What is the combined weight, in pounds, of these 50 tomatoes?

A. 0.0472
B. 11.8
C. 52.36
D. 59
E. 118

Question 2:
The following table gives the times each girl has recorded for seven runnings of the 100-meter run this year. Only one girl may compete in the upcoming tournament. Which girl would you select for the tournament and why?

Race #	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Suzie	15.2	14.8	15.0	14.7	14.3	14.5	14.5
Tanisha	15.8	15.7	15.4	15.0	14.8	14.6	14.5
Dara	15.6	15.5	14.8	15.1	14.5	14.7	14.5

Question 1 is a computation problem (a mathematics problem that requires an answer to a deterministic question) and does not require any statistical thinking or reasoning. To answer it, a student only needs to understand the relationship between a mean and total, presumably found from a formula. The context of the problem is not important – one does not need to consider that there are "prize-winning tomatoes" (Scheaffer, 2006).

Question 2 requires students to use realistic data to investigate a practical question. Unlike Question 1, there is a range of possible answers. Students may use basic mathematical computations (e.g., calculating a mean time for each runner) to help address the question, but to completely answer it, a student needs to use statistical ideas

and thinking, involving the context and nature of variability and their role in this question. Here, making inferences in the presence of variability necessitates knowledge separate from mathematics.

As shown in the two NAEP questions above, statistics and mathematics are used to answer different types of questions. Statistics involves a probabilistic or stochastic way of thinking where inference is supported by evidence, whereas mathematics follows a deterministic style of thinking following the use of a set of axioms to a foregone conclusion supported by an unquestionable chain of logic (Ellis, Bieda, & Knuth, 2012).

The deductive and inductive arguments and methods of teaching are not strictly tied to mathematics and statistics, respectively. A mathematician, for example, may employ both approaches in reasoning through a problem: first, inductive reasoning can be used to form a conjecture about a mathematical object. Then, the mathematician can use deductive reasoning to prove or disprove this conjecture decisively. In this sense, the mathematician can view inductive reasoning as an effective tool for forming conjectures that may then be completely verified through the process of deductive reasoning. In regards to teaching, inductive teaching methods come in many forms, including inquiry-based learning. This style of teaching can be used in both mathematics and statistics classrooms and often involves presenting students with a problem to grapple with, in which they learn what they need to know to address the problem.

With the goal of statistical literacy in mind, both the *SET* and *Pre-K-12 GAISE Reports* recommend methods of teaching that emphasize inductive reasoning (Franklin et al., 2007, 2015). As statistical literacy, reasoning, and thinking are emphasized in

statistics education, teachers first need to demonstrate these skills themselves before attempting to impart them on their students. In consideration of this, researchers have made an attempt to distinguish between these three skills (Ben-Zvi & Garfield, 2004).

Statistical Literacy, Reasoning, and Thinking

The importance of possessing statistical literacy, reasoning, and thinking skills has become more obvious in recent years and especially with the release of the *Pre-K-12 GAISE* and *SET Reports*. The nature of teaching statistics has been evolving from a traditional formula-based approach not emphasizing the aforementioned skills to a data-oriented approach emphasizing conceptual understanding and requiring the acquisition of those skills. Utts (2010) states that after taking an introductory statistics course, students are able to compute basic summary statistics (such as a mean and standard deviation), yet they are often unable to evaluate the misleading statistical results presented daily in the media.

In the article, "What Educated Citizens Should Know about Statistics and Probability" (Utts, 2003), are many examples of poor statistical reports. For example, a headline from a research study published in *USA Today* titled "Prayer can lower blood pressure" (Davis, 1998) is presented. Although there were two groups in this study – one that attended a religious service and one that did not – the participants were not randomly assigned to attend or not attend a religious service. In this case, a cause-and-effect statement such as "prayer can lower blood pressure" should not be made. Yet, this study suggested precisely that in its headline and conclusion, stating

> Attending religious services lowers blood pressure more than tuning into religious TV or radio, a new study says. People who attended a religious service once a week and prayed or studied the Bible once a day were 40% less likely to have high blood pressure than those who don't go to church every week and prayed and studied the Bible less (Davis, 1998).

The headline and conclusion of this study, thus, may mislead citizens into thinking that prayer *causes* lower blood pressure. However, when reading through the description of the study, a citizen who possess statistical literacy, reasoning, and thinking skills would be more likely to recognize fallacious statements in the article (Utts, 2003, 2010). It is important that all students graduate from high school with the skills needed to correctly write statistical results in a coherent manner, correctly interpret statistical results presented to them, and be able to critique and evaluate reports fed to the public on a daily basis (Franklin et al., 2007; Utts, 2003, 2010).

Statistical literacy, reasoning, and thinking are used interchangeably in the literature, and there are no consistent definitions for these three skills. In 2002, the *Journal of Statistics Education* released a sequence of articles focusing on statistical literacy (Rumsey, 2002), statistical reasoning (Garfield, 2002), and statistical thinking (Chance, 2002). In addition, researchers developed a resource to assess students' statistical literacy, reasoning, and thinking skills (Assessment Resource Tools for Improving Statistical Thinking – ARTIST; Garfield, delMas, & Chance, 2006). In light of the advancements and work of these authors surrounding these three skills, I adopt their definitions here.

Statistical Literacy

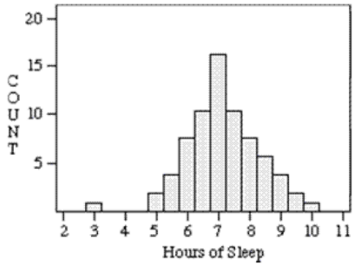
Under its umbrella, statistical literacy includes a number of ideas; more precisely, it entails two outcomes (Rumsey, 2002). First, those who are statistically literate should function as educated members of society; second, they should have a basic understanding of statistical terms and techniques. To achieve fluency in understanding statistics, students need to have a baseline knowledge of its fundamental ideas (Rumsey, 2002; Garfield, delMas, & Chance, 2003; Garfield & Ben-Zvi, 2007). Garfield et al. (2003) define statistical literacy as including

> basic and important skills that may be used in understanding statistical information or research results. These skills include being able to organize data, construct and display tables, and work with different representations of data. Statistical literacy also includes an understanding of concepts, vocabulary, and symbols, and includes an understanding of probability as a measure of uncertainty (p. 7).

Ultimately, statistical literacy revolves around understanding the basic language, concepts, and tools of statistics.

To further emphasize the definition and expectations of statistical literacy, an appropriate example (Garfield et al., 2006) is provided below.

Example 1
The following graph shows a distribution of hours slept last night by a group of college students.



Select the statement below that gives the most complete description of the graph

> in a way that demonstrates an understanding of how to statistically describe and interpret the distribution of a variable.
>
> a. The bars go from 3 to 10, increasing in height to 7, then decreasing to 10. The tallest bar is at 7. There is a gap between three and five.
> b. The distribution is normal, with a mean of about 7 and a standard deviation of about 1.
> c. Most students seem to be getting enough sleep at night, but some students slept more and some slept less. However, one student must have stayed up very late and got very few hours of sleep.
> d. The distribution of hours of sleep is somewhat symmetric and bell-shaped, with an outlier at 3. The typical amount of sleep is about 7 hours and overall range is 7 hours.

Here, a student needs to understand how to describe (shape, center, spread, outliers) and interpret the histogram – a skill of statistical literacy. Many of the ARTIST assessment questions considering statistical literacy ask students to *identify*, *describe*, *translate*, *interpret*, *read*, and *compute* (Garfield et al., 2006).

Statistical Reasoning

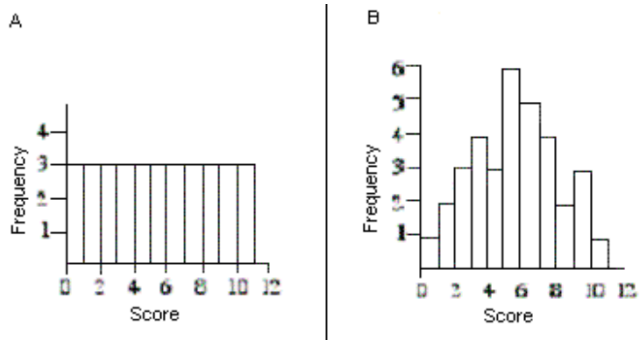
Statistical reasoning involves a person's ability to effectively interact and reason with statistical information, making inferences as appropriate (Garfield, 2002). A conceptual understanding of statistical concepts is vital to possessing statistical reasoning skills (Garfield, 2002). Garfield et al. (2003) define statistical reasoning as

> the way people reason with statistical ideas and make sense of statistical information. This involves making interpretations based on sets of data, representations of data, or statistical summaries of data. Statistical reasoning may involve connecting one concept to another (e.g., center and spread) or may combine ideas about data and chance. Reasoning means understanding and being able to explain statistical processes and being able to fully interpret statistical results (p. 8).

The ARTIST questions assessing statistical reasoning are posed in a way to ask students to explain *why* or *how* – corresponding to the need for a conceptual

understanding to justify solutions and reasoning. An example of a question assessing statistical reasoning is provided below (Garfield et al., 2006).

Example 2
Determine which graph has the higher standard deviation (it is not necessary to do any calculations to answer these questions).



a. A has a larger standard deviation than B.
b. B has a larger standard deviation than A.
c. Both graphs have the same standard deviation.

It is noted that calculations are not needed to answer this type of question; though, based on anecdotal evidence from a similar type of question I provided to a college introductory statistics class, many students *want* to answer this question by using computation. Time is spent on the mere act of computing the two standard deviations rather than harnessing one's conceptual understanding of standard deviation. In short, a student that solely uses calculations to answer this question has demonstrated statistical literacy but *not* statistical reasoning. I could assess their reasoning skills by expanding on the question, prompting students to discuss how they used their knowledge about standard deviation in their answers.

Statistical Thinking

Statistical thinking focuses on the "big picture" of the statistical investigative process (Chance, 2002). One who statistically thinks about a problem knows *when* and *how* to apply statistical tools, knowledge, and procedures, understanding the role of variability in statistics. Garfield et al. (2003) state statistical thinking

> involves an understanding of why and how statistical investigations are conducted and the 'big ideas' that underlie statistical investigations. These ideas include the omnipresent nature of variation and when and how to use appropriate methods of data analysis such as numerical summaries and visual display of data. Statistical thinking involves an understanding of the nature of sampling, how we make inferences from samples to populations, and why designed experiments are needed in order to establish causation. It includes an understanding of how models are used to simulate random phenomena, how data are produced to estimate probabilities, and how, when, and why existing inferential tools can be used to aid an investigative process. Statistical thinking also includes being able to understand and utilize the context of a problem in forming investigations and drawing conclusions, and recognizing and understanding the entire process (from question posing to data collection to choosing analyses to test assumptions, etc.) Finally, statistical thinkers are able to critique and evaluate results of a problem solved or a statistical study (p. 8).

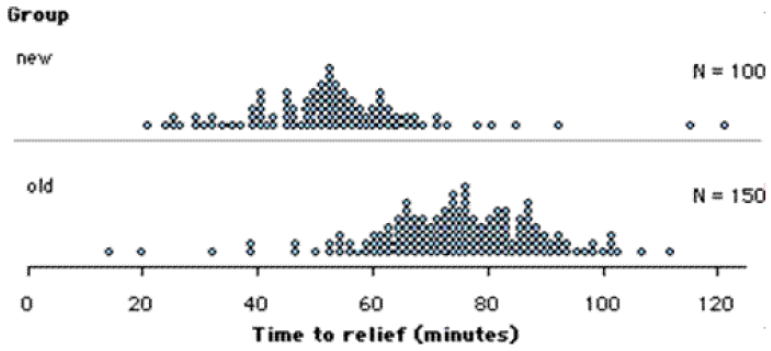
For instance, an individual would need statistical thinking skills to assess the results presented in the *USA Today* study discussed at the beginning of this section. With this skill, an individual knows that by randomly assigning treatments in a study, any lurking variables are "evened out," enabling one to say that any differences observed are indeed caused by the treatment. This individual understands how observational studies and experiments are conducted and the different kinds of inference one can make. With this information, an individual can "critique and evaluate the results" of a statistical study.

Generally, the types of ARTIST assessment questions associated with statistical thinking ask students to *apply*, *critique*, *evaluate*, and *generalize*. For instance, the

example below (Garfield et al., 2006) demonstrates an attempt to assess a student's statistical thinking.

Example 3

A drug company developed a new formula for their headache medication. To test the effectiveness of this new formula, 250 people were randomly selected from a larger population of patients with headaches. 100 of these people were randomly assigned to receive the new formula medication when they had a headache, and the other 150 people received the old formula medication. The time it took, in minutes, for each patient to no longer have a headache was recorded. The results from both of these clinical trials are shown below. For each statement, indicate whether you think the student's conclusion is valid.



a. The old formula works better. Two people who took the old formula felt relief in less than 20 minutes, compared to none who took the new formula. Also, the worst result – near 120 minutes – was with the new formula.
    i. Valid.
    ii. Not valid.

b. The average time for the new formula to relieve a headache is lower than the average time for the old formula. I would conclude that people taking the new formula will tend to feel relief about 20 minutes sooner than those taking the old formula.
    i. Valid.
    ii. Not valid.

To answer this question, students need to use statistical thinking skills to critique the provided conclusions. Appropriately, they need an understanding of the statistical investigative process and how to use the dot plots as an aid in analyzing the data.

The Relationship Between Statistical Literacy, Reasoning, and Thinking

One of the goals of the reform-based movement in statistics education is fluency in statistical literacy, reasoning, and thinking (Garfield & Ben-Zvi, 2004; Franklin et al., 2007, 2015). While these skills are separate, there is some overlap between them (Figure 2; delMas, 2002). Many tasks require a student to use some combination of the three skills. Students need to have a basic understanding of statistical terms, language, and ideas before they are fully capable of explaining, deciding, critiquing, evaluating, and making decisions about information (Rumsey, 2002). That is, students need to be fully literate before they can begin to reason and think statistically. This is immediately apparent in the context of the *USA Today* study. Before an "informed citizen" can critique the results of the study, they must know, for instance, what an observational study is. Students need all three skills to fully understand and be engaged with the statistical investigative process.

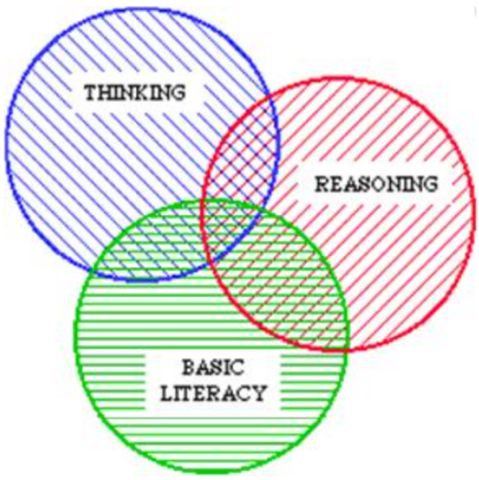


Figure 2. One possible relationship between statistical literacy, reasoning, and thinking (delMas, 2002). The three skills are independent of one another with some overlap.

Statistical Literacy in the *Pre-K-12 GAISE Report*

The *Pre-K-12 GAISE Report* (Franklin et al., 2007) identifies statistical literacy as the "ultimate goal" and states it is "essential in our personal lives as consumers, citizens, and professionals." The report (Franklin et al., 2007) provides examples of what a statistically literate citizen should be capable of doing:

- the statistically literate citizen should understand the behavior of 'random' samples and be able to interpret a 'margin of sampling error' (p. 1);
- a statistically literate person should ask such questions [is the information presented in a scientific finding reliable?] and be able to intelligently answer them (p. 2);
- a statistically literate high-school graduate will be able to understand the conclusions from scientific investigations and offer an informed opinion about the legitimacy of the reported results (p. 2); and
- a statistically literate high-school graduate will know how to interpret the data in the morning newspaper and will ask the right questions about statistical claims (p. 2).

Largely, one who is statistically literate (as defined by the *Pre-K-12 GAISE Report*) understands the language of statistics and can pose meaningful questions, collecting, collating, and analyzing data to investigate them (Bargagliotti et al., 2014; Franklin, 2013).

To become statistically literate (as described in the *Pre-K-12 GAISE Report*), teachers and students must acquire statistical reasoning and thinking skills, with the

understanding that these are grounded in statistical literacy (Figure 3). In this regard, the *Pre-K-12 GAISE Report's* use of "statistical literacy" is intertwined with the definitions of statistical literacy, thinking, and reasoning. In other words, statistical literacy is treated as an "all encompassing goal of instruction" (delMas, 2002). The development of statistical reasoning and thinking skills are goals en route to this larger goal of statistical literacy.

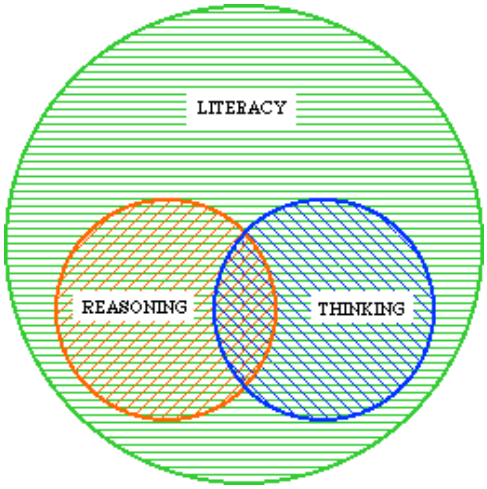


Figure 3. Another possible relationship between statistical literacy, reasoning, and thinking (delMas, 2002). Here, statistical literacy is viewed as the overarching goal of statistics instruction.

The Teaching of Statistics

Despite the increased attention in K-12 statistics education in recent years, teachers of statistics still face many challenges (Franklin et al., 2015; Tishkovskaya & Lancaster, 2012; Batanero et al., 2011). Among these is managing the changing role of statistics in a K-12 mathematics curriculum (Peck et al., 2013). The subject of statistics is new for many of these teachers (Franklin et al., 2007), and with the adoption of national

and state content standards, high school teachers must now teach statistics in non-AP Statistics classes (Franklin et al., 2015). There are often high expectations for these teachers, though many are not familiar enough with statistics to properly teach it (Franklin et al., 2015; Franklin & Kader, 2010). Teachers, often as a result of their own education, hold a limited conception of statistics as a collection of formulas and procedures (Peck et al., 2013).

The changes in what is now expected for the teaching of statistics do not just revolve around the increased scope of statistics concepts; there are shifting demands in the approach needed to teach these concepts, as well. The recommended approach is now more data-oriented than formulaic, and this is made possible with the advancement of technology (Franklin, 2013; Batanero et al., 2011). Students in modern classrooms are able to uncover important concepts with hands-on activities that take advantage of technology (Franklin, 2000). Accordingly, high school teachers must develop an "understanding of statistical reasoning from a data and simulation perspective and an appreciation for the effectiveness of such an approach in teaching the basic tenets of statistics" (Franklin et al., 2015).

As statistics is a relatively new subject in grades K-12, it is not surprising that many in-service teachers are not prepared to teach it (Lazar & Franklin, 2015; Batanero et al., 2011) or to use the appropriate technological tools at the level suggested by the *Pre-K-12 GAISE Report, SET Report* and *CCSSM* (Franklin et al., 2015). This is due, in part, to most in-service teachers being trained prior to the adoption and implementation of national and state statistics content standards. Moreover, they have often not been

exposed to an accurate treatment of the underlying principles of data analysis that they must teach (Franklin et al., 2007). Therefore, it is necessary for many teachers to become familiar with statistical concepts, ideas, and technology not taught when they were students (Pratt, Davies, & Connor, 2011).

The Culture of K-12 Statistics

The integration of statistics in the K-12 mathematics curriculum is difficult, as it is still not a widely-held belief that, as a subset of mathematics, statistics and data analysis are important (Lazar & Franklin, 2015; Batanero et al., 2011). The culture in high school mathematics is that, for many students, knowledge of calculus is ideal; however, in a world surrounded by data, statistics often should take this place instead (Lazar & Franklin, 2015). Still, teachers and administrators do not generally feel an urgent need to teach statistical content over mathematical content, and in turn develop students' statistical literacy, reasoning, and thinking skills.

Assessments often drive what is taught and how it is taught in the classroom (Bargagliotti & Franklin, 2015). There is a lack of statistics content standards assessed in state and national assessments; those that include questions categorized in the realm of statistics often focus on procedures and computations rather than conceptual understanding (Franklin et al., 2015). Lazar and Franklin (2015) note that until assessments begin to include important statistical ideas, educators will not consider statistics itself an important subject.

Teachers' beliefs about statistics also greatly influence the manner in which they teach statistics and how comfortable they are in teaching it (Batanero et al., 2011). A

common belief among K-12 teachers is that they do not have a foundational level of statistics knowledge and are not equipped to teach statistics (Franklin, 2000, 2014). They also believe their students have a harder time understanding statistics than mathematics, and that, as teachers, they are not well-equipped to overcome these difficulties (Batanero et al., 2011). Many in-service mathematics teachers lack confidence about delivering a data-driven curriculum and being able to teach and interpret the statistics content standards (Lazar & Franklin, 2015; Franklin, 2013). The nature of data and randomness feels foreign, as does the extensive focus on communication in statistics (Garfield & Ben-Zvi, 2008).

With statistics being taught under the umbrella of the K-12 mathematics curriculum, teachers' beliefs about instructional goals for mathematics will also ultimately influence how they teach statistics (Batanero et al., 2011). Some teachers, for instance, may be partial to mathematical applications and modeling over formal mathematics; accordingly, they may teach statistics differently. Their instructional strategies in mathematics may influence their decision to use a formula-based approach rather than a data-oriented approach to teach statistics. Furthermore, teachers often draw from their own experiences as students. For example, teachers who were not taught with technology may not be comfortable using it themselves in their own teaching (Batanero et al., 2011).

K-12 Teachers' Statistical Content Knowledge

A given teacher's statistical content knowledge depends on a variety of factors, and many teachers do not have a solid enough grasp on the material to teach to the

requirements of the curriculum (Batanero et al., 2011). Most in-service K-12 mathematics teachers have little to no experience in statistics, as teacher preparation programs have typically required few statistics courses. Those that have been exposed to statistics often were taught with a traditional, formula-based approach or received a limited amount of it (as footnotes of a book primarily devoted to mathematics) (Franklin, 2000, 2013; Bargagliotti, 2014). Decidedly, teachers, at no fault of their own, often lack statistical content knowledge.

A teacher's statistical content knowledge directly impacts his or her effectiveness while teaching, as it influences the decisions he or she makes when teaching statistical concepts (Batanero et al., 2011). Franklin (2013) states one of the major concerns she hears from teachers revolves around their lack of statistical content and pedagogical knowledge. This creates difficulty in asking "good" questions and effectively answering student questions, and these teachers fear being "put on the spot" by students (Franklin & Kader, 2010).

At the Joint Mathematics Meetings in 2010, Allan Rossman stressed the importance of being able to ask students "good" questions (Peck et al., 2013). To teach statistics effectively, teachers' understanding of statistical concepts must be considerably more sophisticated than that of their students (Peck et al., 2013). Without this understanding, teachers may not be comfortable to solicit or ask "good" questions of their students. "Good" statistics questions incorporate meaningful context that is relevant and includes the ideas of variability, conceptual understanding, and interpretation (Peck et al., 2013). The book, *Developing Essential Understanding of Statistics: Grades 9-12* (Peck et

al., 2013), provides examples of statistics questions ranging from "not very good" to "good."

To begin with, a teacher might display a dataset of 10 numbers and pose the following question:

> **Question 1:**
> "What is the value of the standard deviation of the following 10 numbers?"

To answer this question, students only need to demonstrate computational skills; in addition, there is no context to these numbers that needs to be addressed. A teacher could add context to this question (e.g., "The numbers in this dataset represent the lifetime, in hours, of 10 light bulbs from a new brand that your school is considering for use in the football stadium light fixtures" (Peck et al., 2013)), but this information is irrelevant for what students are asked to do with the data – they still simply need to compute a standard deviation. An even "better" question to ask is

> **Question 2:**
> "The standard deviation for the lifetimes of bulbs from the brand currently in use is 40 hours. What does the standard deviation that you computed for the sample of light bulbs from the new brand tell you about how this brand might compare with the old brand?"

Here, students need to move beyond computation and be able to address the context of the problem, communicating a deeper understanding of standard deviation. Teachers who are more confident in their statistical content knowledge are more likely to move beyond asking students to perform simple computations.

Another issue that may arise from a lack of statistical content knowledge is a passing-on of common misconceptions. For example, if a teacher does not fully understand the construction or message of a sampling distribution, his or her students

may not come to fully understand these ideas either (Batanero et al., 2011). Many teachers themselves have not experienced learning statistical concepts through simulation or the four-step process of statistical investigations (as prescribed by the *Pre-K-12 GAISE Report*) making it difficult for them to fully implement these ideas in their own classrooms.

The Role of Technology

The face of mathematics and statistics instruction – for students and teachers alike – has been greatly changed by technology (Lee & Hollebrands, 2011; Chance et al., 2007; Batanero et al., 2011). Class sets of computer-based technologies are becoming more readily seen in the K-12 classroom, and the call to use technology in the mathematics classroom is omnipresent (e.g., NGA, 2010a, Franklin et al., 2015). Many organizations are currently promoting the use of technology and have described its importance in learning. In their *Principles to Action Report*, NCTM (2014) states "an excellent mathematics program integrates the use of mathematical tools and technology as essential resources to help students learn and make sense of mathematical ideas, reason mathematically, and communicate their mathematical thinking." The Association of Mathematics Teacher Educators' (AMTE) 2015 position statement on technology supports the NCTM's position stating,

> technology can be used to facilitate mathematical discovery, reasoning, and connections that may be difficult or impossible without its use. Activities that engage students in connecting multiple representations, and those that invite students to analyze or create images, visualizations, and simulations provide wide-ranging opportunities for mathematical exploration and sense-making.

Additionally, the *Pre-K-12* and *College GAISE Reports*, *SET Report*, *CCSSM*, and the 2008 Joint ICMI/IASE Study Conference call for teachers to use technological tools, including simulations of random processes.

Technology has changed the way statisticians work and, therefore, how statistics is taught in schools (Chance et al., 2007). Students have the opportunity (with technology) to become active learners and *do* statistics rather than focus on the use of formulas. Technology provides sophisticated tools for teaching and using simulations that enable students to understand difficult statistical mechanisms without a lifetime of mathematical study (Franklin, 2013). The efficiency of technology frees up time in class to focus on interpreting and understanding key statistical concepts – time that was spent on tedious efforts to compute means or construct exploratory plots and visualizations of data (Franklin, 2013). For instance, one can forgo the use of statistical tables (e.g., *z*-tables) to look up an approximate *p*-value and use a technological tool that can produce more accurate results much more quickly (Chance et al., 2007). Technology has and continues to make a difference in statistics education and has a positive impact on statistics education if used appropriately (Chance et al., 2007; Franklin, 2013).

Technology can be used as a productivity tool (for efficiency and automation) and/or a cognitive tool (for conceptual understanding). On one hand, technology allows one to carry out more calculations with higher accuracy, producing visuals near-instantly. It expedites computation that could be completed by hand at a much slower rate (Lee & Hollebrands, 2011; Chance et al., 2007; Pratt et al., 2011). However, the *effective* use of technology does not solely reinforce the emphasis on computation; instead, students

should use it in a way that places statistical literacy, thinking, and reasoning at the forefront of learning (Pratt et al., 2011).

Technological tools can help students cognitively, affording them an opportunity to acquire statistical knowledge and develop a conceptual understanding of statistical concepts. Technology (when used cognitively) engages students in *doing* and understanding statistics (Lee & Hollebrands, 2011; Chance et al., 2007). For example, simulations can be used as a pedagogical tool to enhance the conceptual understanding of statistical inference, sampling distributions, and confidence intervals (Chance et al., 2007). In this vein, technology is used in ways that extend what one may be able to do without it, helping to deepen one's statistical understanding. Cognitive technologies, however, do not replace the teacher's role in the classroom, as it is still necessary for teachers to observe students, address any difficulties, and answer questions (Chance et al., 2007).

Technology, of course, can be used as *both* a productivity and a cognitive tool. Sometimes, using technology in a productive capacity to quickly compute a summary statistic (e.g., standard deviation) or produce multiple graphical representations of data (e.g., boxplots and histograms) can reduce cognitive load in this area, so that instructional focus can now be placed on the "bigger picture" of statistics (Lazar & Franklin, 2015; Lee & Hollebrands, 2011; Chance et al., 2007). For example, a calculator or computer can quickly compute a standard deviation, but the larger advantage is that students are now able to devote more attention to interpreting and understanding the standard deviation, ultimately deepening their concept of variation.

In statistics, there are many different types of technology that can be used. Chance et al. (2007) break down these types into seven categories: (1) statistical software packages, (2) educational software, (3) spreadsheets, (4) web applets/stand alone applications, (5) graphing calculators, (6) multimedia materials, and (7) data and materials repositories. Statistical software packages, often used by statisticians (e.g., **R**, SAS, Minitab), are developed with performing statistical analyses in mind and thus have full capabilities to do such analyses. Educational software packages (e.g., TinkerPlots), on the other hand, have limited statistical capabilities and are not as powerful as full-fledged statistical software packages. These tend to be more user-friendly and less overwhelming for students. Spreadsheets such as Excel can be used for statistical computations, tests, and visual displays. However, one must use caution and check how the computation algorithms are defined; in many instances it can be quite cumbersome to create an appropriate plot. The amount of web applets available has grown tremendously in recent years. This type of technology often helps students explore concepts in a visual and dynamic environment. Graphing calculators are a great portable tool that mathematics students typically have access to. Still, beyond statistics courses, they are not generally used by statisticians as a substitute for statistical software packages. Multimedia materials often seek to combine several types of technology (e.g., videos, applets, statistical software packages). Finally, data and material repositories are locations of real datasets and engaging activities to be used with students.

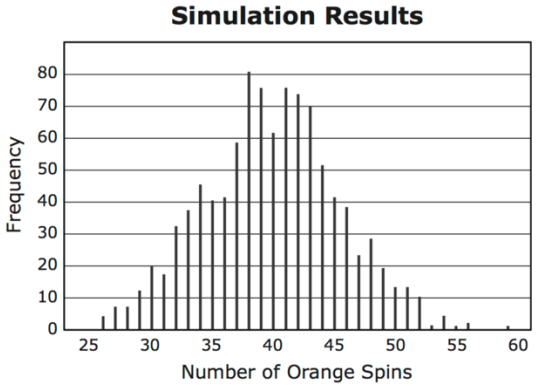
While there are multiple types of technology out there to choose from, technology should not just be used merely for the sake of using it. Teachers must decide which is the

most appropriate for engaging their students in the learning process (Chance et al., 2007). The effective uses of technology in the statistics classroom can be summarized as "automation of calculations" (to free up time to spend on conceptual understanding), "emphasis on data exploration," "visualization of abstract concepts," "simulations as pedagogical tool," and "investigation of real life problems and datasets" (Chance et al., 2007). In particular, the use of simulations to study random processes and statistical concepts is becoming more and more popular in the statistics classroom (Franklin, 2013; Chance et al., 2007). Concepts such as sampling distributions, confidence intervals, and patterns in randomness can become clearer with the use of simulations, and students are afforded an opportunity to explain things that would not be as clear if they were reading it on a page of a textbook (Chance et al., 2007; Lazar & Franklin, 2015; Lee & Hollebrands, 2011).

Guidelines for the use of simulations have been appearing in the *CCSSM*, *Pre-K-12 GAISE Report,* and *SET Report*. For example, the *CCSSM* states students are to build intuition about inference as they "generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions." In response to the adoption of the *CCSSM*, the Partnership for Assessment of Readiness for College and Careers (PARCC) Assessment has been developed to measure students' knowledge surrounding the *CCSSM*. One of their sample assessment questions surrounds the idea of using results from a simulation.

> **PARCC Question:**
> A circular spinner is divided into five sectors of different colors. A student spun the arrow on the spinner 200 times and recorded that the arrow stopped on the orange sector 38 times out of the 200 spins. To test whether the spinner was fair,

the student used a computer to simulate the number of times the arrow stops on orange in 200 spins of a fair spinner equally divided into five sectors of different colors. The results of 1,000 trials of the simulation are shown.



Based on the results of the simulation, is there statistical evidence that the spinner is not fair?

a. Yes, because 38 was the most frequent outcome.
b. Yes, because about 8% of the outcomes were 38.
c. No, because the distribution is approximately normal.
d. No, because an outcome of 38 or less is not unusual.

Based on these guidelines and assessment questions, there is a need for teachers to effectively use technology and incorporate simulations when teaching statistical concepts.

Using technology, however, is obviously not without its challenges for both teachers and students. Many teachers need specific training to become familiar with technology and integrating it into the classroom in ways that support inquiry-based learning, engage students, and promote conceptual understanding. In practice, teachers have to make many decisions about how to effectively integrate technology in their classrooms. These decisions are influenced by their content knowledge, understanding of how to teach and how students learn, and understanding of how to use and apply technology in their teaching.

Knowledge and Teaching

Teaching is a multifaceted skill that uses a wide range of knowledge. In 1986, Shulman argued research and practice on teaching should not exclusively focus on either content or pedagogy, as one in the absence of the other does not entirely represent a teacher's knowledge. He suggested a third major component of teaching, called *pedagogical content knowledge* (PCK), overlapping both content knowledge and pedagogical knowledge; this is a type of knowledge specific to teaching, representing the way in which teachers combine their pedagogical knowledge with their content knowledge (Shulman, 1986). His idea, going beyond the consideration of content and pedagogy as isolated entities and instead blending these together, has been an influential contribution in teacher education research. To implement instructional techniques that maximize a student's potential to learn content, teachers must, accordingly, fully understand the effect content and pedagogy have on each other.

In 1987, building on his earlier work, Shulman developed the model of *Pedagogical Reasoning and Action* (PRA) encompassing the six actions a teacher ought to take during teaching. This emphasizes teaching as "comprehension and reasoning, as transformation and reflection." Teaching, he argues, beings with some "text" (e.g., lesson plan, textbook), and a teacher is immediately challenged by the task of synthesizing what he or she already knows to enable it to be taught (Shulman, 1987). PRA cycles through the actions of comprehension of subject matter, transformation of subject matter, instruction, evaluation, and reflection.

Technological Pedagogical Content Knowledge

The use of technology in the classroom adds depth and complexity to instruction (Koehler & Mishra, 2005; Mishra & Koehler, 2008). While technology has emerged as a prominent force in education, changes in the process of teaching and learning have also occurred. It is not enough for teachers to simply learn how to use current types of technology, as technology inevitably becomes obsolete (Koehler & Mishra, 2005; Hollebrands & Zbiek, 2004). Teachers must also develop new techniques and skills to teach with a set of technological tools that are constantly in flux. With that, researchers also need new ways of perceiving the phenomena of integrating technology in the classroom. Extending Shulman's idea to technology (that is, technological knowledge alone is not enough to effectively integrate technology in the classroom), many researchers have created terms to understand a teacher's knowledge about effectively integrating technology in the classroom (e.g., Koehler & Mishra, 2005; Pierson, 2001; Angeli & Valanides, 2005; Niess, 2005; Margerum-Leys & Marx, 2002).

Koehler and Mishra (2005) introduced the term *technological pedagogical content knowledge* (TPACK), a framework developed to describe the scope of a teacher's knowledge in the process of effectively integrating technology into his or her teaching. In their view, teacher knowledge involves many relationships between content, pedagogy, and technology (Koehler & Mishra, 2005; Mishra & Koehler, 2006). They argued the effective integration of technology requires consideration of all three components and the many relationships between them (Mishra & Koehler, 2006).

Koehler and Mishra (2005) were not the first to suggest that effective teaching with technology requires an understanding of multiple domains of knowledge. Pierson (2001) used TPCK (analogous to TPACK) to define a teacher's skill in the integration of technology in his or her classroom. Similar terms were also created by researchers: examples of this include information and communication technology (ICT)-related PCK (Angeli & Valanides, 2005), technology-enhanced PCK (Niess, 2005), and PCK of educational technology (Margerum-Leys & Marx, 2002). However, Mishra and Koehler (2006) argue that what sets their framework apart from others is the consideration of pedagogy, content, and technology *and* their articulation of the relationships between these three components. They go beyond seeing these three components by themselves, instead stressing the relationships between them. Aside from looking at each component separately: *pedagogical knowledge* (PK), *technological knowledge* (TK), and *content knowledge* (CK), there is a need to look at them in pairs and all together: *pedagogical content knowledge* (PCK), *technological content knowledge* (TCK), *technological pedagogical knowledge* (TPK), and finally *technological pedagogical content knowledge* (TPACK). Thus, their TPACK framework consists of seven components.

The Components of TPACK

In their work, Mishra and Koehler describe the seven components of their TPACK framework (Koehler & Mishra, 2005; Mishra & Koehler, 2006, 2008). *Content knowledge* (CK), simply put, is knowledge of the various concepts and theorems central to the subject matter to be taught or learned. *Pedagogical knowledge* (PK) broadly encapsulates knowledge unique to methods of teaching. This includes, but is not limited

to, knowledge regarding the management of one's classroom, planning lessons, and better enabling students to perform to their potential in class. *Technological knowledge* (TK) refers to knowledge of how to use types of technology such as presentation tools and more advanced examples, such as computers. More importantly, TK covers the ability to adapt to and learn new technologies. *Technological content knowledge* (TCK) is knowledge about the interlocking nature of technology and content. Teachers need to be able to identify the optimal technologies for a given subject matter. *Technological pedagogical knowledge* (TPK) is knowledge about the interrelation between technology and pedagogy. Teachers excelling in this category are able to understand how a given technology may enhance a student's learning potential, while also adapting their respective teaching styles to accommodate this technology. *Pedagogical content knowledge* (PCK) reflects Shulman's idea that effective teaching requires more than separate understanding of content and pedagogy; different content is more adequately suited to different methods of teaching. Finally, *technological pedagogical content knowledge* (TPACK) is the basis of effective teaching with technology – this is a form of knowledge that extends past CK, PK, and TK alone, including but ultimately overshadowing the three components alone. It requires a deep understanding of the use of technology in teaching various concepts, placing some emphasis on the effective use of appropriate pedagogical techniques. Further, it is necessary to be able to identify the roadblocks students face in truly understanding a given content area.

TPACK Framework in Mathematics

The TPACK framework proposed by Koehler and Mishra (2005) is general and not tied to one specific subject. Niess (2005) focused her efforts on TPACK in *mathematics* and researched what it means to prepare pre-service mathematics teachers to teach with technology. Niess (2005) delineated four components in teachers' knowledge in integrating technology:

1. (Conception): "An overarching conception about the purposes for incorporating technology in teaching subject matter topics." This includes teachers' knowledge of their own subject and what students ought to learn, as well as the role of existing technologies in learning.
2. (Students): "Knowledge of students' understandings, thinking, and learning in subject matter topics with technology." Teachers are wholly dependent on their own beliefs regarding students' understandings and how these are affected by technology.
3. (Curriculum): "Knowledge of curriculum and curricular materials that integrate technology in learning and teaching subject matter topics." Teachers contemplate and integrate technology for particular topics, and they need to understand how concepts can be arranged in a technology-based curriculum.
4. (Strategies): "Knowledge of instructional strategies and representations for teaching and learning subject matter topics with technologies." Teachers

alter their lessons for specific technologies, and they use particular representations with technologies to satisfy the needs of their classes.

Over the course of several years, she and other researchers began to focus on understanding how a teachers' TPACK in mathematics develops and how they progress to acquire TPACK. In 2007, Niess, Sadri, and Lee proposed a developmental model for TPACK in mathematics, stemming from Everett Roger's (1995) model of the innovation-decision process. The idea described therein is that it is not the case that teachers either do or do not have TPACK; rather, there are varying levels of TPACK. Their development model is based on a five step decision-making process in which an individual decides to accept or reject a given technological innovation. These five stages include the following levels.

1. Recognizing (knowledge): teachers are able to use technology and see where it applies but do not effectively implement technology in their classroom.
2. Accepting (persuasion): teachers form a deep-seated opinion regarding the role of technology in their instruction.
3. Adapting (decision): teachers put themselves in a position where they are compelled to either use or reject the use of technology.
4. Exploring (implementation): teachers accept the role of technology in their classrooms and begin to effectively implement it.
5. Advancing (confirmation): teachers reflect upon the implications of coupling their teaching with technology.

AMTE supported this developmental model and has long advocated integrating technology into the mathematics classroom and preparing teachers to effectively teach with it. In this vein, and to help consider these various levels of TPACK, AMTE created a visual representing Niess et al.'s (2007) developmental model (Figure 4). This visual depicts the levels teachers progress through as they develop TPACK, combining technology, pedagogy, and content (Niess et al., 2009). It is the case that, as teachers progress throughout the steps of the model and grow their technological knowledge from their existent pedagogical content knowledge, TPACK will emerge (Niess et al., 2009).

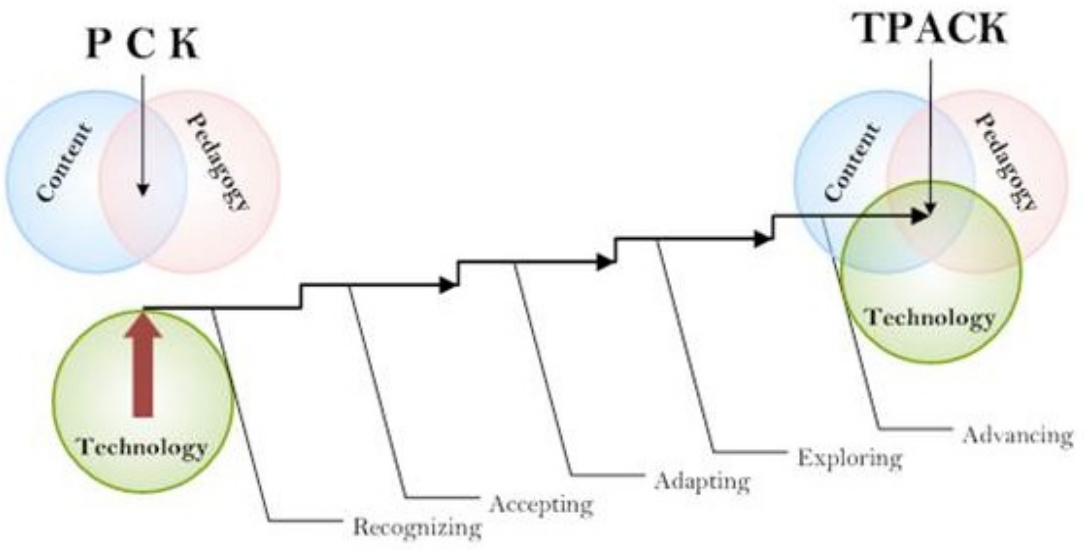


Figure 4. AMTE's visual of Niess et al.'s (2007) mathematics teachers' TPACK developmental model.

TPACK Framework in Statistics

Although TPACK is not subject-specific, TPACK in mathematics may not be the same as TPACK in statistics. To fill this void, researchers have conceived some structure for TPACK in statistics (Sorto & Lesser, 2010; Lee & Hollebrands, 2011; Lesser &

Groth, 2008). Accordingly, the term *technological pedagogical statistical knowledge* (TPSK) has been established, along with a general framework devised by Lee and Hollebrands (2011) and Lesser and Groth (2008) (Figures 5 and 6, respectively). Lee and Hollebrands (2011) argue TPSK emerges from *statistical knowledge* (SK) and *technological statistical knowledge* (TSK). Lesser and Groth (2008), conversely, entertain a perspective analogous to Mishra and Koehler's (2005) in which TPSK is where technological knowledge, statistical content knowledge and pedagogical knowledge collide.

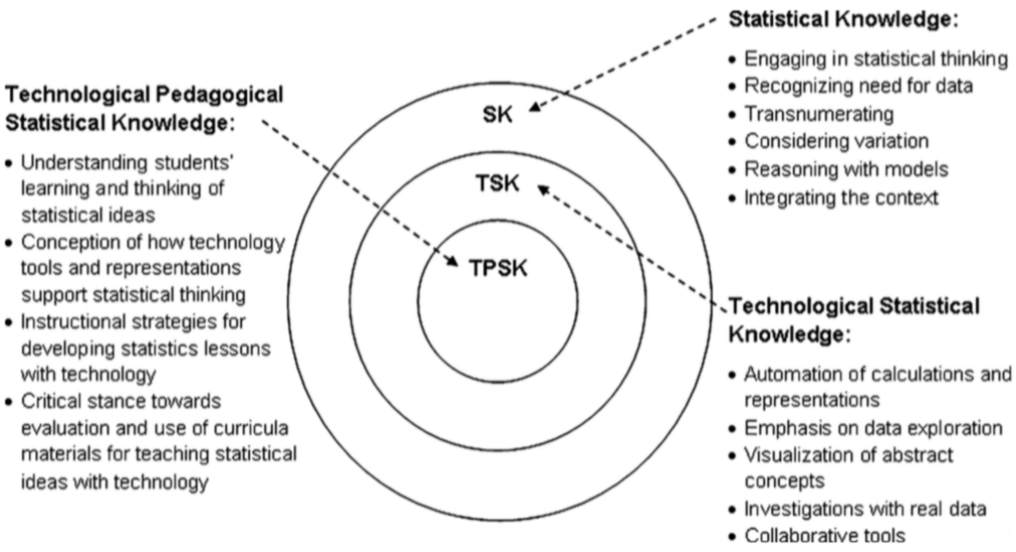


Figure 5. Lee and Hollebrands' (2011) interpretation of a *Technological Pedagogical Statistical Knowledge Framework.*

While Figures 5 and 6 exhibit some differences, the underlying concepts they are describing are, more or less, the same framework in a different context. Their messages are clear: the separated knowledge of content, pedagogy, and technology is not sufficient to effectively teach statistics with technology. Rather, one must acknowledge the multiple

types of knowledge (content, pedagogical, and technological), together with their many connections with one another.

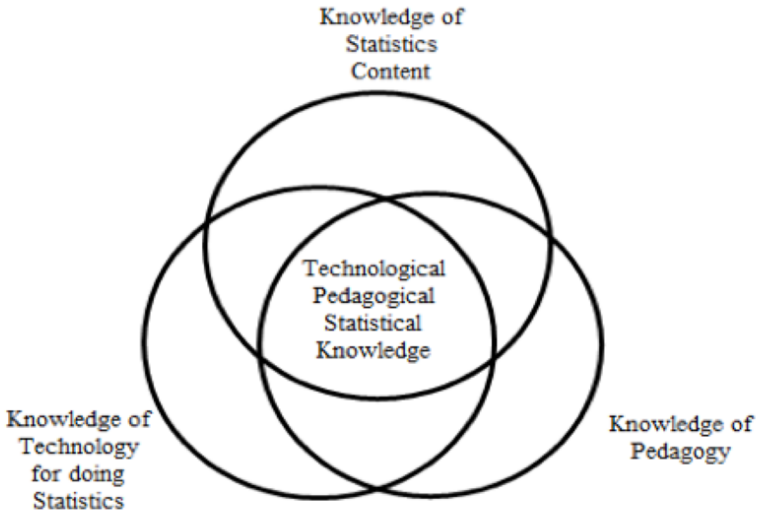


Figure 6. Lesser and Groth's (2008) interpretation of a *Technological Pedagogical Statistical Knowledge Framework*.

Measuring TPACK

The TPACK framework (Koehler & Mishra, 2005) and development model (Niess et al., 2007) can serve as a conceptual lens through which researchers view the development of teachers' TPACK, and their readiness for effectively incorporating technology in the classroom. In light of this, numerous instruments and measurement techniques supporting the TPACK framework and development model have been established throughout the years. A teacher's TPACK is often measured through a combination of several of these instruments (Voogt, Fisser, Roblin, Tondeur, & van Braak, 2012). While over 100 TPACK instruments and measurement techniques have

been developed (Koehler, Shin, & Mishra, 2011), they can generally be classified in four broad categories: surveys, design tasks, teaching observations, and interviews.

Surveys are self-report measures in which participants are asked to rate the degree to which they agree with a given statement. These statements often relate to the seven components of knowledge (e.g., CK, PCK, TCK, etc.) under the umbrella of the TPACK framework and are one of the most popular methods used to measure teachers' proficiency within each component. For example, Schmidt et al. (2009) created and validated a 47-item *Survey of Preservice Teachers' Knowledge of Teaching and Technology* to measure elementary pre-service teachers' perceptions of their own TPACK understanding. A flaw inherent in any self-report measure lies in the participants' limited ability for self-assessment (Abbitt, 2011). The survey results may very well misestimate teachers' TPACK in either direction, and so while they can be reliable in identifying shifts in teachers' perceptions of self, they should not be fully trusted to assess the degree to which teachers understand the components of TPACK (Brantley-Dias & Ertmer, 2013).

Another popular method of measuring TPACK is the use of performance assessments or design tasks. These are typically used for evaluating teachers' TPACK by directly examining teacher-created artifacts. Among the possible artifacts examined are lesson plans or other design tasks mimicking real tasks a teacher would perform in his or her classroom. By peering into the design and planning process, researchers can directly assess a teacher's TPACK. Some TPACK performance assessments ask teachers to create portfolios, lesson plans, and reflective journals, whereas others include instructional

scenarios and cases that necessitate deeper levels of TPACK to solve (Koehler et al., 2011; Brantley-Dias & Ertmer, 2013; Abbitt, 2011). To analyze these results, researchers may use qualitative or quantitative measures, with the former being typically used in classifying responses with emergent themes. In addition, a participant's entries may be evaluated based on rubrics lying within the TPACK framework (e.g., Tournaki & Lyublinskaya, 2014). Ultimately, the chief concern with performance assessments is a given teacher's ability to communicate their understanding of TPACK; it is tricky to document understanding of TPACK if a teacher's ability to communicate this understanding is inadequate.

An additional way to measure TPACK is through observations of teacher-enacted lessons. These observations often encompass video recording or field note-taking (for the purpose of examining how a teacher's TPACK is transferred into the classroom). The observations are typically analyzed qualitatively through a conceptual lens inspired by a theoretical framework described by the researcher. A relevant drawback with these observations, however, is that a researcher is only able to get a "snapshot" of a teacher's TPACK at any given time. Thus, multiple observations of the same teacher are required to obtain a more complete picture of how a teacher's knowledge about integrating technology is transferred into practice.

Finally, interviews may be composed of a set of oral questions asked by the interviewer of the teacher. Usually recorded with audio- or video-tapes, interviews are transcribed, coded, and then analyzed to measure teachers' TPACK, and interviewers may ask follow-up questions as needed. Interviews tend to be analyzed from a qualitative

perspective, with these interviews providing valuable insight into the depth of teachers' TPACK. Some limitations include the potential reluctance of teachers to express themselves or them responding in ways they believe the interviewer wants.

Improving Teaching

Teaching can be improved; however, there is a misconception that a student's poor performance is the result of the inherent qualities of teachers (Hiebert & Morris, 2012). As documented by Stigler and Thompson (2009) and Hiebert and Morris (2009, 2012), the primary focus to improve teaching in the U.S. has been to (1) recruit more talented people and retain them longer, and (2) to improve the qualifications of both pre- and in-service teachers. Both of these approaches predominately focus on improving the *characteristics* (or *qualities)* of teachers. While more-skilled and knowledgeable instructors may be better equipped to use some methods of teaching, it is unclear as to whether or not the methods themselves are preferable to those used by a more average teacher (Stigler & Thompson, 2009). Moreover, while individual teachers can grow to become more effective instructors through professional development opportunities, these gains are essentially lost as the current generation retires and the next generation begins teaching (Hiebert & Morris, 2009, 2012; Morris & Hiebert, 2011). Addressing these concerns, Hiebert and Morris (2012) argue for an alternative method to improve teaching in their idea that teaching can be improved through the development of a base of knowledge specific to teaching. This alternative approach emphasizes the methods and

tools used in teaching rather than those teaching with them and may create long-lasting and continual improvement over generations of teachers (Stigler and Thompson, 2009).

The core of this alternative approach focuses on developing, sustaining, refining, and storing a shared knowledge base for teaching where teachers can take advantage of what others have learned before them (Hiebert & Morris, 2009, 2012). In this sense, new teachers do not have to "start over." Still, there is the problem of transmission. For this approach, it is necessary for the relevant knowledge to be stored and accessible for the next generation of teachers; it needs to be described effectively enough so individuals other than those who devised it can implement it similarly; and finally, a base of teacher experiences regarding the success or failure of their implementation needs to be established to measure the degree to which this knowledge was transmitted. If these conditions can be met, then it may be possible for the quality of teaching, regardless of teachers' experience, to improve over time (Hiebert & Morris, 2009, 2012; Stigler & Thompson, 2009).

Working on the methods used to teach can take the form of artifacts or instructional products (Hiebert & Morris, 2009, 2012; Morris & Hiebert, 2011). These are public materials for teaching that can be improved over time as they are created and shared, implemented by teachers, and then edited to reflect the realities and issues of implementation. These are seen as having the most potential to improve long-term instruction (Stigler & Thompson, 2009), as they simultaneously provide a baseline that guides actual instruction while allowing for a store of knowledge regarding implementation in practice.

Specially annotated lesson plans are one type of artifact that may be particularly useful and form an ever-expanding store of teaching knowledge (Hiebert & Morris, 2012; Morris & Hiebert 2011). To be shareable, these lesson plans are well-detailed and complete – to the point where they constitute a curriculum in and of themselves (Hiebert & Morris, 2012; Morris & Hiebert, 2011). In general, annotated lesson plans contain knowledge about "what to do" and "why and how to do it that way" (Hiebert & Morris, 2012). They exhibit a number of characteristics designed to maximize student achievement and promote ease of implementation in the classroom. This information is included, in enough detail, so that all teachers may, regardless of skill level, implement lessons as an expert instructor would. First, learning goals for each particular lesson are established in a clear and concise way to facilitate the appropriate choice and precise implementation of a given lesson. Next, reasons for important instructional decisions are laid out so that lessons are adaptable without undermining the spirit of the lesson. Further, to anticipate a students' thinking during class – so teachers can anticipate and use their thoughts for a more useful and authentic lesson – students' potential responses to various tasks and prompts are supplied. Finally, generally useful information such as summaries of key concepts, student misconceptions, and hints to avoid mistakes is provided (Hiebert & Morris, 2012; Morris & Hiebert, 2011).

Morris and Hiebert (2011) observe disparities in teachers' backgrounds which may lead them to implement the same lesson plan in different ways. If the goal is consistent, high-quality instruction, this is indeed a concern, but one that is remedied (at least in part) by the use of annotations. Again, annotated lesson plans are theoretically

designed with enough detail to equalize the effectiveness of teachers' implementations, regardless of expertise. Variance in implementation can be caused by a number of factors, such as the way students respond to activities. Hence, while variation will of course persist to some extent, the annotated lesson plans can be continually revised to limit variation as these factors come to light (Morris & Hiebert, 2011).

Theoretical Framework

Qualitative research relies on the use of a theoretical framework – a conceptual lens through which researchers view data and inform the research (Creswell, 2013). In this regard, the TPACK framework (Koehler and Mishra, 2005) and developmental model (Niess et al., 2007) allow me to establish what aspects of teacher knowledge are key to integrating technology in statistics while also providing a common basis for conversing about this study's findings.

Central to the notion of teaching effectively with technology is the role of inquiry-based teaching. The idea that education should not just be a transmission of ideas and processes from teacher to student is not new (Dewey, 1910). Over the years, the composition of mathematics curricula has changed, becoming more broadly standards-based. With this, a shift from preparing students for a standardized test toward the development of students' critical thinking and reasoning skills has occurred. Marked by the adoption of the *CCSSM*, and with recommendations included in the *Pre-K-12 GAISE* and *SET Reports,* statistics content standards have become more focused, stressing the importance of conceptual understanding, and statistical literacy, reasoning, and thinking.

This affords teachers and students an opportunity to delve deeper into the statistics content.

In stark contrast to its position in the era of high-stakes testing, student inquiry has come to prominence during the standards-based movement (Barrow, 2006). This strategy, intertwined with the history of constructivism, has been a part of education and is currently one strategy used to implement the *CCSSM* content standards and mathematical practices. Inquiry-based learning was recommended by John Dewey (1910) in K-12 science education. He believed teachers should do more than tell students relevant information; students, he argued, must be prompted to actively explore the material in order to learn.

In short, inquiry is a learning process emerging from the constructivist theory of how students learn, in which students construct their own knowledge by learning and reflecting on their own experiences. Inquiry-based learning, therefore, can provide opportunities for students to improve their own conceptual understanding of material via investigation. Rather than merely stating an established fact, teachers pose questions allowing for students to explore, investigate, and engage in discussion as they seek their own path to an answer. The teacher presents lessons that are student-centered and engaging, guiding students to build upon previous knowledge and construct meaning; in this context, the student is an active learner rather than a passive receiver of information.

Inquiry is not, however, a “free-for-all” that allows for students to focus on whatever they please; it is a collaborative, investigative, engaging, and intellectual experience (Barrow, 2006). It is necessary for the teacher to make the experience

purposeful. Teachers do lecture, but inquiry-based learning also requires time for students to pursue the answers to their own questions and make some decisions related to their learning.

This research focused on the effective use of technology in the classroom. In other words, using cognitive technologies to develop students' conceptual understanding of statistics in conjunction with inquiry-based learning. As such, the choice of constructivism, inquiry, and the TPACK framework and developmental model as a theoretical framework to examine the data in this study was natural.

Summary

Through this literature review, I have described how the K-12 statistics curriculum has evolved from a formula-based approach (focused on computation and procedure) to a data-oriented approach (focusing on conceptually understanding statistical concepts). This evolution in purpose has been broadly supported by many educational organizations in their national releases promoting statistical literacy as a goal for all citizens. Ultimately, teaching and learning statistics requires a different style of thinking that current K-12 mathematics teachers may not be acclimated to, creating a pronounced need to help these teachers improve their instruction of statistics, particularly with technology.

The alternative approach to improving teaching through the improvement of the methods teachers use, as proposed by Hiebert and Morris (2012), focuses on creating, implementing, and refining transmissible and shareable artifacts. This research explored

the use of annotated lesson plans as a means to improving teaching statistics with technology. As statistics content standards continue to be implemented at the K-12 level, it is vital to have well-reviewed and effective resources guiding in-service teachers to effectively teach statistics with technology.

CHAPTER 3

METHODOLOGY

Introduction

The field of statistics education has evolved tremendously over the past two decades. There is now a growing motivation to shift from a traditional formula-based style of teaching statistics to a more data-oriented approach emphasizing conceptual understanding and statistical literacy. To implement this type of approach in the classroom, it is necessary to effectively integrate technology into the teaching of statistics. Statistics, however, is a relatively new subject for most in-service K-12 mathematics teachers (Franklin et al., 2007), and many need training in using technology (U.S. Department of Education, 2016). This discrepancy highlights the need to help foster and develop in-service K-12 mathematics teachers' ability to effectively use technology when teaching statistics. In this vein, annotated lesson plans may be a means to centralize specific knowledge and guide the instruction of statistics with technology. The general goal of this study, therefore, was to investigate the extent to which knowledge of teaching statistics with technology is passed along to teachers via annotated lesson plans. From this goal, three specific research questions emerged and were refined throughout the course of this study.

1. To what extent do specially annotated lesson plans enable in-service high school teachers to effectively integrate technology in their instruction of statistical concepts?

2. Do the qualities inherent in specially annotated lesson plans carry over into in-service high school teachers' implementation of a future technology-assisted statistics lesson?
3. How are the following qualities associated with an in-service high school teacher's use of specially annotated lesson plans as a means to guide the instruction of statistics with technology?
    a) The teacher's statistical content knowledge;
    b) the teacher's awareness of the use of inquiry in a mathematics classroom;
    c) the teacher's awareness of the use of technology in a mathematics classroom; and
    d) how the teacher's implementation of a specially annotated lesson plan aligned with its goals.

In the remainder of this chapter, I clarify my (the researcher's) position toward using technology in the classroom and provide additional background information to identify my biases and ability to minimize these biases when collecting and analyzing data. Next, I provide a detailed description of the research design, including information about the setting, participants, and instrumentation. Finally, I discuss the methods used in collecting data, and the style of analysis (qualitative and quantitative) used to investigate the research questions.

The Researcher

Prior to entering the Ph.D. program in Mathematics with a specialization in Mathematics Education, I received a Bachelor of Arts in Mathematics, followed by a Master's of Science in Mathematical Modeling. A certificate in faculty preparation at the college level accompanied the latter. During my dissertation research (focusing on statistics), I also completed a Master's of Science in Statistics.

I have taught a variety of college level mathematics and statistics courses, including Pre-Algebra, Elementary Algebra, Intermediate Algebra, College Algebra, Calculus II, Number and Operations for K-8 Teachers, Higher Mathematics for K-8 Teachers, and Introductory Statistics (taught solely using a traditional as well as a data-oriented approach). As part of my Ph.D. program, I also completed an internship at a local high school, where I spent eight weeks primarily in two different AP Statistics classrooms. I devoted one week of the internship to teach high school students how to use GeoGebra to analyze their data and present results. During this internship, I also occasionally attended two Algebra II classes.

My implicit research biases are apparent from my experiences as a student and teacher. Generally, technology has been at the forefront of most of my educational experience. As an undergraduate and graduate student of mathematics and statistics, I was able to use technology for visualizing and understanding key concepts. As an instructor, being solely responsible for the instruction of many mathematics and statistics courses, I make the decision as to how to integrate technology into my courses. Technology, when used effectively, can greatly enhance students' learning in

mathematics and statistics; tedious computations that take an exceedingly long time to compute by hand, for instance, can be done in seconds with technology, allowing for a nuanced discussion of context and concept over numbers and calculations.

The classes I have taught form a wide body of curricula and approaches; some have affirmed the traditional formula-based approach, while others have eschewed this in favor of inquiry-based learning. I have taught a Technology-Enhanced Active Learning (TEAL) version of a college introductory statistics course twice and a traditional version of the same course once. The former can be described as an active learning classroom, relying heavily on the use of technological tools, real data, and simulation to teach. In this environment, I witnessed firsthand how technology and simulations can positively impact students' depth of understanding of statistical concepts.

There is, of course, more to the use of technology than simply double-checking calculations. Technology opens up a broad world that enables students to explore and discover mathematical and statistical concepts that are ordinarily inaccessible. Hence, based on my experiences and beliefs, I immediately recognize my bias towards the use of technology in a mathematics and statistics classroom.

Technology ought to be a key part of any classroom that promotes statistical literacy. Still, regardless of my preference towards implementing technology in my own courses, I recognize that it presents a challenge for many teachers. To counteract my bias, it was necessary for me to take care in ascertaining how teachers use particular technologies in their classes. No explicit judgment therein was made, as I concentrated solely on how the technology itself was being used.

The Research Design

In this study, I employed quantitative and qualitative measures and methods of analysis, including the development and use of a completely randomized block design to investigate changes (if any) in in-service high school teachers' use of technology when teaching statistics (Figure 7). To begin with, I established the population of interest; I followed this with recruiting and selecting participants for the study. Next, I sorted those selected participants into homogeneous groups based on their statistical content knowledge, awareness of the use of technology, and awareness of the use of inquiry in a mathematics classroom. I then randomly assigned participants to receive an *annotated* statistics unit (treatment) or a *non-annotated* statistics unit (control). I observed participants implement four statistics lessons from the unit, using an adapted TPACK rubric to evaluate each participant's integration of technology and inquiry in his or her classroom. Here, I recorded multiple "technology scores" for each participant for each of the four lessons taught. The quantitative outcomes of interest were the change in these scores from Lesson 2 to Lesson 5 and the raw scores from Lesson 4. Upon completion of the classroom observations, I conducted interviews and gave participants two final surveys to complete. I discuss these individual components in more detail below.

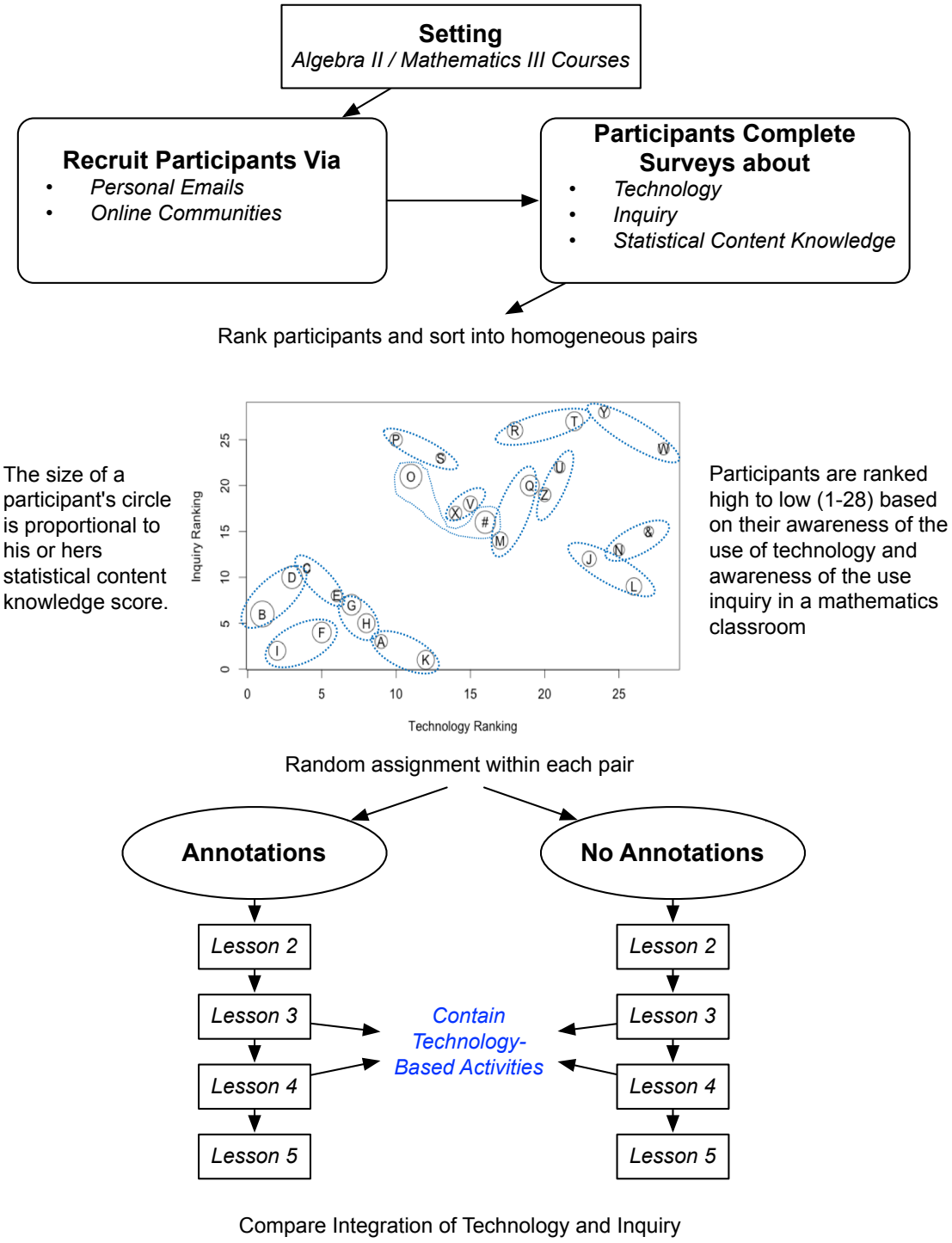


Figure 7. The experimental design I developed for this study.

Population of Interest

The *CCSSM* organizes the high school content standards into six conceptual categories: number and quantity, algebra, functions, geometry, modeling, and statistics and probability (NGA, 2010a). For the purpose of organizing these standards into courses, Achieve, Inc. (a group partnered with the writers of Common Core) developed "*Model Course Pathways in Mathematics*" (NGA, 2010b). These pathways describe ways to bundle the *CCSSM* content standards into rigorous courses and are particularly useful for states and school districts attempting to accomplish this (NGA, 2010b).

The two pathways I considered in this study are the "traditional" and "integrated" pathways (Figure 8). These pathways consist of a sequence of three core courses (each of which include statistics content standards) before taking a higher level course, such as Pre-Calculus, Calculus I, Advanced Statistics, or Discrete Mathematics. The traditional approach, more common in the U.S., is colloquially referred to as the "geometry sandwich." Here, high school students take Algebra I, followed by Geometry and then Algebra II. The second approach, more common in international programs though not completely absent in the U.S., is the integrated approach; this consists of a sequence of three sequential levels of Mathematics I, II, and III.

Both pathways contain a unit, in Algebra II and Mathematics III, devoted to "*Inferences and Conclusions from Data*," where

> students see how the visual displays and summary statistics they learned in earlier grades relate to different types of data and to probability distributions. They identify different ways of collecting data – including sample surveys, experiments, and simulations – and the role that randomness and careful design play in the conclusions that can be drawn (p. 36; NGA, 2010b).

In particular, there are six statistics content standards to be taught in this unit (Table 1).

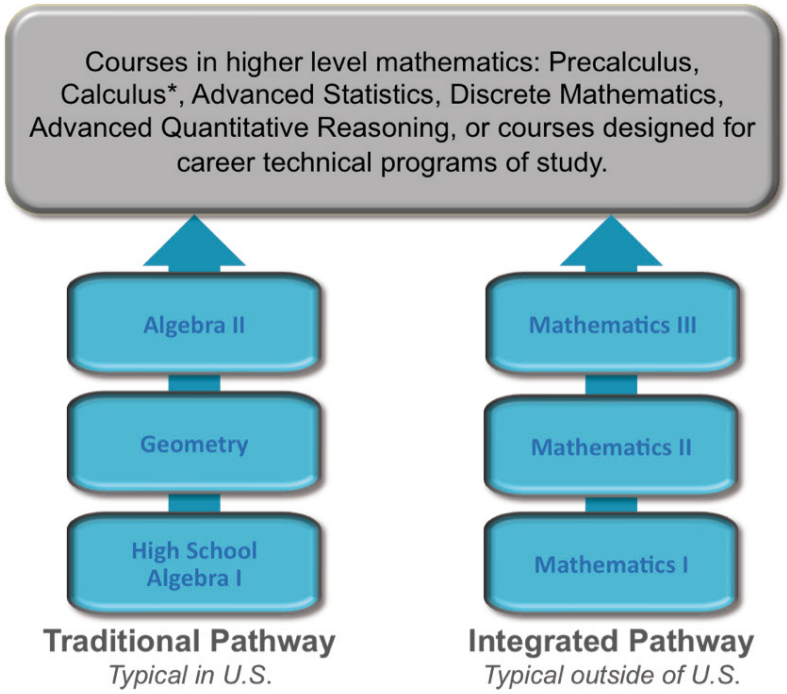


Figure 8. The traditional and integrated course pathways suggested by the *CCSSM* (NGA, 2010b).

These content standards provide statistics content areas in Algebra II and Mathematics III for which technology is well-suited to aid teaching and offer opportunities for students to engage with the material. Illustrating this point, three of them specifically identify the use of simulation, and Franklin (2013) discusses how technology provides sophisticated tools for using simulation. As this study focuses on the effective use of technology in teaching high school statistics, the setting is Algebra II, Mathematics III, or the equivalent classroom in public high schools. Furthermore, it was necessary for students in these classes to have access to a computer or tablet because the activities provided to teachers in this study require Internet access. Therefore, the target

population was in-service public high school teachers in the U.S. who are teaching Algebra II, Mathematics III, or the equivalent course and whose students have access to an Internet-enabled computer or tablet.

Table 1. A list and description of the statistics content standards recommended to be taught in an "*Inferences and Conclusions from Data*" unit within Algebra II and Mathematics III (NGA, 2010b).

Content Standard	**Description**
S.ID.4	Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under a normal curve.
S.IC.1	Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
S.IC.2	Decide if a specified model is consistent with results from a given data-generating process, e.g., using **simulation**.
S.IC.3	Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
S.IC.4	Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of **simulation** models for random sampling.
S.IC.5	Use data from a randomized experiment to compare two treatments; use **simulations** to decide if differences between parameters are significant.

The Statistics Units

I developed two statistics units (annotated and non-annotated) for this research study (Appendix A); the material in these units covers all the statistics content standards recommended for an "*Inference and Conclusions from Data*" unit in Algebra II and Mathematics III (Table 1). The development and distribution of a statistics unit (aligned with the *CCSSM*) also provided an incentive for in-service high school teachers to

participate in this study. For instance, one teacher expressed interest from the start of the recruitment phase, as statistics is a current topic of discussion at his high school (specifically within Algebra II classes). In this case, the teacher and school were discussing how to fit statistics content standards in their curriculum, and he mentioned how the statistics unit could help with their situation.

Each unit contains five lesson plans and the topics of each lesson are displayed in Table 2. The lesson plans within the two statistics units were identical in almost all respects – with the exception of annotations in Lessons 3 and 4 within the annotated statistics unit.

Table 2. The topic(s) of each lesson within the two statistics units.

Lesson	**Topic(s)**
1	Randomness
2	Normal Distributions
3	Observational Studies and Experiments
4	Interval Estimation and Margin of Error
5	Comparing Two Treatments from a Randomized Experiment

The five lessons contain the following items: a detailed lesson plan (which includes annotations is Lessons 3 and 4 within the annotated statistics unit), activities and solutions (except for Lessons 2 and 5), and sample homework questions and solutions. The lesson plans follow a standard template, first displaying the topic(s) of interest, a summary of the lesson, activities to use, estimated time of implementation, pre-requisite knowledge, learning objectives, content standard(s) addressed, assessments, and materials needed. The lesson plans include a detailed outline of "teacher activities," including a

description of the concepts to be taught and recommended examples to work through and statistical definitions to cover.

Lessons 3 and 4 are the only two lessons that contain activities to use with students; these activities focus on using technology and simulations to explore statistical concepts and develop conceptual understanding. The activities contain approximately 15 questions and first guide students to conduct a simulation by hand (with cards) and then make use of a web applet to conduct the simulation with technology. The descriptions and questions within the activities are meant to guide students to explore with technology and think statistically about various concepts.

The lesson plans within these units are not intended to force scripted classroom sessions. Rather, enough detailed information is included to allow participants to obtain a sense of the concepts to be taught. I informed participants that they could implement the lessons in any way they see fit, and if they did not agree with something in a lesson, they could choose to overlook it or alter their teaching approach. For instance, in Lesson 2, I provided various examples involving the normal distribution. These were categorized into two types of problems: finding probabilities and normal percentiles. Many participants chose to overlook the latter, focusing only on finding probabilities.

I did not develop the lessons from scratch. I collaborated with multiple instructors of a college introductory statistics course to revise the lessons and activities used in their courses. The goal of instruction in this college course aligns with one that emphasizes conceptual understanding through an inquiry-based implementation involving realistic data and technology. Consequently, many of the lessons and activities in this course

involve the use of technology and simulation. These lessons and activities are based on the National Science Foundation-funded CATALST (Change Agents for Teaching and Learning Statistics) project (Garfield, delMas, & Zieffler, 2008), which emphasizes this approach.

Because the lessons were originally created for a college introductory statistics course, I adapted them for high school. To do this, I consulted with local in-service high school teachers regarding the content and length of the lessons. This gave me additional knowledge in how to adapt the lessons to the appropriate time length for high school classes. Particularly, some questions in the activities used in the college course were shortened, simplified, or removed, and some descriptions within these activities were also revised. Finally, the context of each lesson was revised to revolve around a common theme: music – something high school students have some genuine interest in. I implemented parts of the lessons in a college setting and further refined the statistics units before disseminating them to participants in January 2016.

The statistics units also contain possible summative assessments (and solutions). These include exam questions, along with a description of a poster project; the project affords students an opportunity to apply their knowledge learned throughout the unit to a real-world context. In the poster project, students will formulate a question of interest, collect and analyze appropriate data, and interpret and communicate their results in a poster.

Participants

Eleven in-service public high school teachers, meeting the stated criteria, volunteered to participate and agreed to implement a statistics unit. These eleven participants were teaching in the U.S. in western, eastern, and Rocky Mountain states. No two participants were teaching in the same school district. They had a wide range of teaching experiences, educational backgrounds, statistical content knowledge, and awareness of the use of technology and inquiry in a mathematics classroom. While their awareness of the use of inquiry and exploration ranged from non-existent to developed, they all used technology at varying levels in their mathematics classrooms.

Recruiting and Selecting Participants. Participants were recruited from May 2015 to December 2015, with invitations sent out in two waves to online communities of mathematics teachers and to personal e-mails of accessible public high school teachers in two comparable rural states. These states were chosen out of convenience because of personal connections to teachers in them. Online communities were preferred, as they connect a variety of teachers with one another to provide for professional learning. The inherent ease of communicating over the Internet bridges the gap between a teacher's desire for continued learning and his or her time constraints, and this is immediately apparent in the existence of many online communities (like GeoGebra Institutes and Park City Mathematics Institute Groups) and social networks (Twitter, Facebook, Google Pages) that provide, among other things, assistance for teachers seeking lessons on a particular topic or news in their respective fields. The teachers in these communities have already self-identified with the desire to learn more about mathematics and statistics, and

they may already use or share resources and ideas disseminated by these respective communities.

When I began to recruit participants from online communities, I created a GeoGebraBook (akin to a webpage). Here, I provided information about myself, described the context of my research study, and included a link to the first of two recruitment surveys. This GeoGebraBook can be viewed at (https://www.geogebra.org/m/cDFN5XSh?doneurl=%2Farnold52).

Recruitment surveys came in two phases. First, an enrollment survey prompted interested teachers for their names, work e-mail addresses and home states, and to identify whether or not they would be teaching Algebra II, Mathematics III, or an equivalent course during the 2015-2016 academic school year. Overall, there were 70 high school teachers interested who were teaching the appropriate course. Next, I sent these 70 teaches an initial survey consisting of four parts: a statement of informed consent (Appendix B), basic information about the individual's workplace and accessibility to a class set of computers or tablets, current awareness about the use of technology and inquiry in a mathematics classroom, and a statistical content knowledge assessment. Upon completion of the initial survey, I provided each teacher a $25 Amazon gift card.

Of these 70 interested teachers, 28 had access to a class set of computers or tablets, completed the initial survey, signed the consent form, and volunteered to participate. I invited all 28 to participate in the study and receive a statistics unit. By the end of December 2015, 11 were still receptive to this. Those who chose not to participate

in the study primarily cited concerns with time constraints, being observed, or conflicts with administrative decisions. For instance, one teacher who declined to participate stated

> I have to teach 'the book, the whole book, and nothing but the book.' I know the kids NEED the statistics unit; however, the powers that be state that I cannot add anything that will slow down the completion of the book.

I chose my deadline for recruiting and accepting participants to be the end of December 2015 as opposed to the beginning of the school year in August 2015, as one of the confirmed participants was planning to begin teaching the statistics unit in January 2016. Furthermore, all others who volunteered to participate stated they would not teach the statistics unit until March 2016 at the earliest.

Treatment Allocation. Once in-service public high school teachers were recruited, were selected, and agreed to participate, I established how to randomly allocate the treatments (annotated or non-annotated statistics unit) using an iterative process on the 28 teachers who originally opted in. To begin with, I considered a completely randomized design (Figure 9) in which each participant would be randomly assigned to receive the annotated or non-annotated statistics unit. Here, each participant has the same chance of receiving the annotated statistics unit.

This design is one of the simpler experimental designs. In theory, the two resulting groups – those receiving annotations and those not receiving annotations – are, on average, similar. This type of design works well when experimental units (e.g., high school teachers) are fairly homogeneous and no obvious confounding variables can be identified. However, teachers vary in characteristics such as educational background, teaching experiences, content knowledge, teaching philosophies, instructional strategies,

and so on. As a result, I considered how I could restrict the randomization to account for as many of these sources of variability as possible in the design. Accounting for them in the design allows the comparison of technology scores from those who received the annotations to those who did not to be made between maximally similar teachers; this helps ensure treatment differences are not confounded with differences in other teacher characteristics.

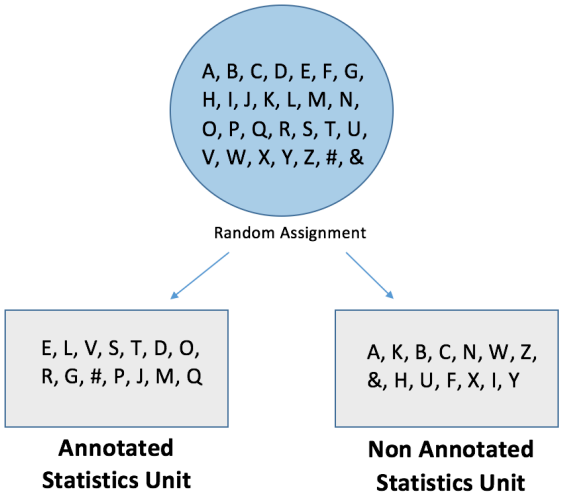


Figure 9. A diagram of a completely randomized design in which participants would be randomly assigned to receive the annotated or non-annotated statistics unit.

In an effort to account for the varying characteristics of teachers, I first considered sorting participants into groups with similar characteristics (blocks). Within each block, participants would then be randomly assigned to receive the annotated or non-annotated statistics unit. This strategy of blocking results in an experimental design controlling for known sources of variability among study participants that are not of direct interest to the researcher but can impact the results of the study. In a sense, blocking "evens out the playing field" by creating more homogeneous groups within which the treatments are

randomly assigned. When effective, blocking helps control for negative effects of confounding variables and extra sources of background variability before making treatment comparisons, thus, making it easier to detect treatment effects if they exist.

To create the blocks, I identified three "teacher characteristics" that would have the largest impact on the way participants responded to the treatment: (1) awareness of the use of technology, (2) awareness of the use of inquiry, and (3) statistical content knowledge. A teacher's awareness of the use of technology and inquiry in their mathematics classroom influences the instructional decisions they make. For example, if a teacher believes technology is primarily a productivity tool, he or she will likely use technology for computational efficiency in his or her classroom. On the other hand, if a teacher believes technology is primarily a cognitive tool, he or she will likely use technology to support inquiry-based learning. Additionally, because statistics is a relatively new subject for many teachers, I expected high school teachers to vary in their statistical content knowledge, impacting their ability to correctly convey statistical concepts to their students.

I first asked participants six open-ended questions (in the initial survey) to qualitatively assess (and rank) their awareness of the use of technology and awareness of the use of inquiry in a mathematics classroom. In addition, I asked each participant to complete a short assessment (in the initial survey) to quantitatively assess their statistical content knowledge. I used these responses to sort participants into blocks; however, there are many different ways one can create reasonable blocks based on these three variables. I first considered blocking by a teacher's awareness of the use of technology and

awareness of the use of inquiry in a mathematics classroom. I developed four potential blocks (low inquiry/low technology, low inquiry/high technology, high inquiry/low technology, and high inquiry/high technology) and sorted participants into these four blocks (Figure 10).

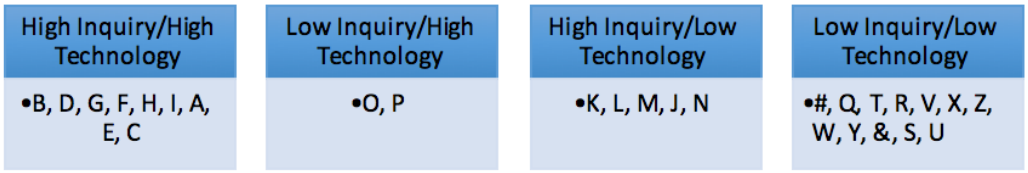


Figure 10. Sorting participants into four blocks based on a teacher's awareness of the use of technology and awareness of the use of inquiry in a mathematics classroom.

However, the cutoffs between low and high technology and low and high inquiry are arbitrary, and this approach does not consider the participants' statistical content knowledge. When one uses this particular blocking scheme and then randomly assigns treatment conditions to participants within each block, by chance, all participants with high statistical content knowledge could be those receiving the annotations, adversely impacting the usefulness of the results. I, therefore, wanted to also account for each participant's statistical content knowledge. Thus, I considered choosing arbitrary cutoffs (low, medium, high) for the statistical content knowledge scores. This created multiple blocks based on technology, inquiry, *and* statistical content knowledge (Figure 11). Within each block, I then would randomly assign participants to receive the annotated or non-annotated statistics unit.

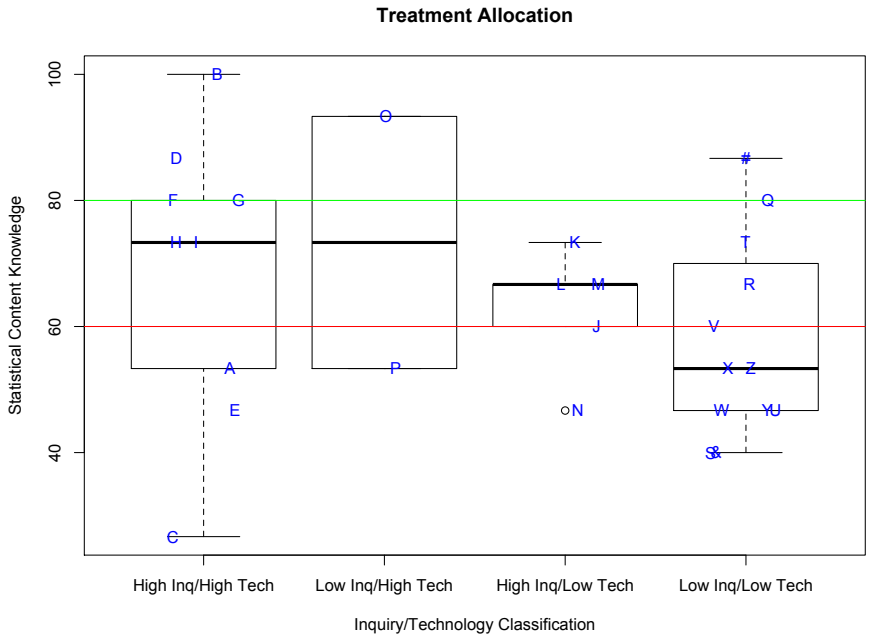
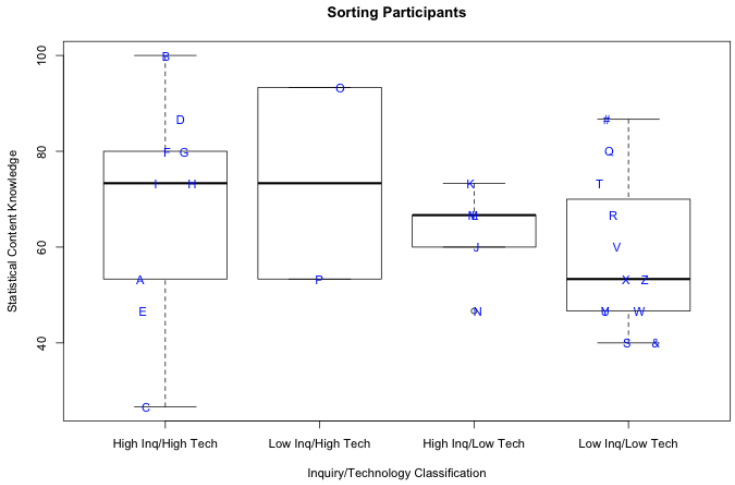


Figure 11. Blocking by technology, inquiry, *and* statistical content knowledge. I created arbitrary cutoffs for the statistical content knowledge scores.

Yet, when I created these cutoff values for the numeric statistical content knowledge scores, I noticed I was discretizing these scores and losing additional information about participants. Furthermore, at the time of this study, I did not find any guidelines on how to classify teachers' statistical content knowledge into blocks, making any choice of cutoff values completely arbitrary. I thus considered a different method; I would keep the four blocks described above (based solely on the levels technology and inquiry) and then, within each block, pair participants based on their statistical content knowledge scores, randomly assigning annotations to the halves of each pair (Figure 12).



Pairs	
B, D	N, ---
G, F	#, Q
H, I	T, R
A, E	V, X
C, ---	Z, W
O, P	Y, U
K, L	&, S
M, J	

Figure 12. Blocking by technology and inquiry, and pairing participants (within each block) based on their statistical content knowledge scores.

This strategy indeed allowed me to account for statistical content knowledge in my design. Furthermore, pairing participants (rather than creating groups of four to five individuals) reduces as much subject-to-subject variability as possible before making treatment comparisons. Therefore, participants within each group are as similar as they can be, and any ultimate differences within the pairs in terms of the outcome are more likely due to the treatment they received.

However, I found that in using this strategy I was not thoughtfully considering those participants at the endpoints of each block (i.e., the boxplots in Figure 12). For example, a participant may be sorted into the high inquiry/high technology block despite being on the low end of the scale within this block (e.g., Participant "C" in Figure 12); this would, of course, possibly make the participant more comparable with someone from another group (like Participant "K" in the high inquiry/low technology block in Figure 12). I concluded this method of blocking and pairing participants was too coarse – the

inherent problem being the three teacher characteristics were treated as discrete as opposed to continuous phenomena.

Treatment allocation based on continuous covariates is another, though less common, strategy similar to blocking based on discrete variables (e.g., Hu & Hu, 2012). In an effort to treat a teacher's awareness of the use of technology and awareness of the use of inquiry more continuously, I *ranked* participants on both these characteristics. Then, using these two rankings, along with the raw statistical content knowledge scores, I created a bubble plot and sorted participants into as similar of pairs as possible, randomly assigning participants within each pair to receive the annotated or non-annotated statistics unit (Figure 13). This strategy resulted in a more thoughtful blocking strategy that still considered all three teacher characteristics. This process of ranking participants, creating a bubble plot, and pairing participants is described in more detail later, in the section "Analyzing the Data."

Participants had no knowledge of whether they received the annotated or non-annotated statistics unit; only I was aware of this. Accordingly, this was a single-blind experiment. Participants were invited to one of two Google Drive Folders (containing either the annotated or non-annotated statistics unit); here they were only able to anonymously view the contents in each folder. In addition, the names of these folders concealed the fact that one unit contains annotations and the other does not. The annotated statistics unit was in the folder named "Statistics Unit" while the non-annotated unit was in the folder named "Stat Unit." Careful attention was made to ensure each

folder contained the correct lessons and that I invited the correct participants to the appropriate folders.

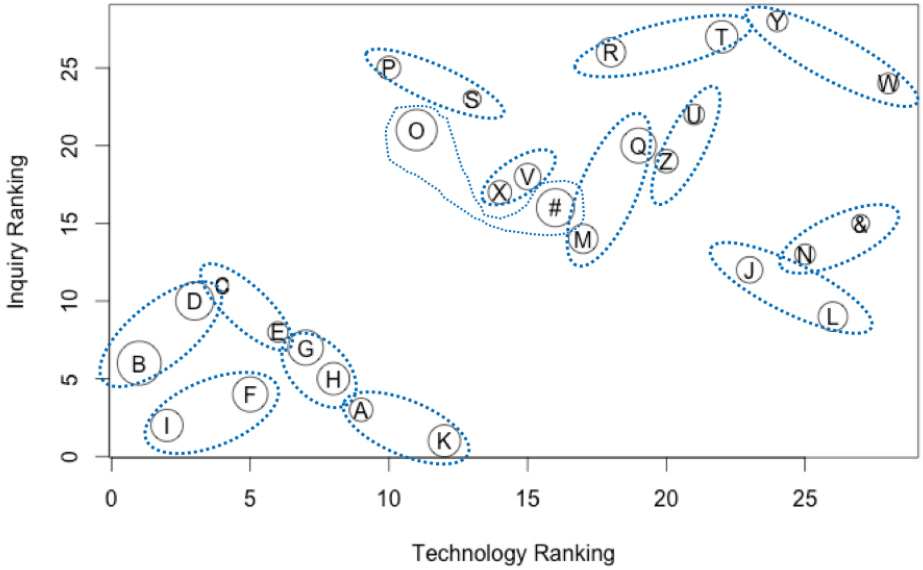


Figure 13. Using the technology and inquiry rankings, along with the raw statistical content knowledge scores to sort participants into as similar of pairs as possible. Here, the size of an individual's circle corresponds to his or hers statistical content knowledge score.

I chose to blind participants to the unit they received and withhold the goals of the study to minimize any bias. Participants may have responded differently to the units if they knew one contained annotations or may have used technology in a manner differently than they normally would. For example, participants who received the annotated statistics unit may have implemented the lessons differently than if they had received the non-annotated statistics unit, ultimately impacting the results of the study.

TPACK Rubric

To assess changes in how teachers integrate technology into their instruction, Lyublinskaya and Tournaki (2012) initially developed a "TPACK Levels Rubric" specific to the TI-Nspire, and then later adapted their rubric to encompass other instructional technologies (Tournaki & Lyublinskaya, 2014). Their rubric is structured as a matrix, with each cell representing a specific level of TPACK (identified by Niess et al., 2007) in each of the four components of teachers' knowledge in integrating technology identified by Niess (2005). Within each cell are two performance indicators; the first focuses on how teachers and students are using technology, while the second focuses on the digital, technology-based materials accompanying the lesson. The possible score in each of the four components (rows) is an integer or half-integer between 0 and 5; whole numbers represent both performance indicators have been met, while a half-integer represents only one being met (Lyublinskaya and Tournaki, 2012; Tournaki & Lyublinskaya, 2014). A numerical score of 0 denotes the participant did not meet the standards of the lowest level (recognizing). A teacher's overall TPACK level, as defined by Tournaki and Lyublinskaya (2014), is the lowest score across all four components. In this research study, I used an adapted TPACK rubric (Appendix C) to evaluate the integration of technology and inquiry in a lesson.

The process of adapting the rubric and testing it occurred throughout January 2015 to March 2016. Because the original TPACK Levels Rubric was created for *mathematics*, I made a conscious effort to observe *statistics* classes to evaluate how well the rubric applied to the teaching and learning of statistics. During the Spring 2015 and

Fall 2015 semesters, I observed (directly and through video-recorded observations) and evaluated several instructors teaching a college introductory statistics course with the original TPACK rubric. During this time, I made notes about what did and did not fit within the scope of the rubric and what was observable in a *recorded* classroom observation. In addition, one of my research participants (who began teaching the statistics unit in January 2016) taught three sections of Algebra II and agreed to record each of her classes during the implementation of Lessons 2 and 3. With another researcher, I randomly selected and used two of these sets of classroom observations to test my adapted rubric. This was possible because the remainder of my participants did not begin implementing the unit until March 2016. I used the third set of classroom observations for "officially" assessing this participant's use of technology and inquiry when teaching statistics.

When examining Tournaki and Lyublinskaya's (2014) TPACK Levels Rubric, one may observe a clear continuum as the levels of recognizing, accepting, adapting, exploring, and advancing are met (Figure 14). There is an apparent shift from the teacher being the sole user of technology as a productivity tool to the students primarily using technology to explore mathematical concepts and develop conceptual understanding.

The two lowest levels of the rubric (recognizing and accepting) are centered *teachers* primarily using technology as more of a productivity tool than a cognitive tool. At these levels, technology is used for motivation, concentrating on procedures and skills. For instance, a teacher may demonstrate how to use the graphing calculator to find the five-number summary statistics of a dataset. Here, a teacher would show students how to

enter the data in the calculator and press the correct sequence of buttons to display the summary statistics; students may or may not follow along with their own calculator. In this situation, there is no context or interpretation of the numbers explained, and students are not provided with opportunities to explore and develop a deeper understanding of concepts.

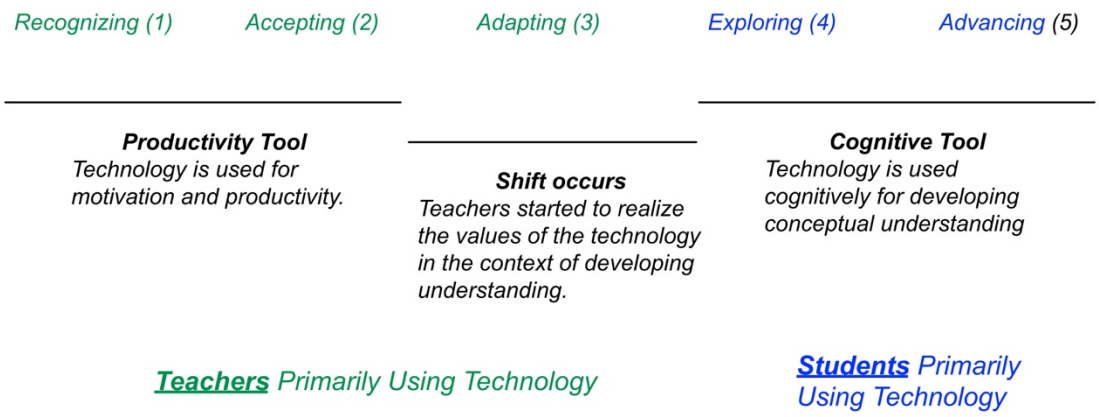


Figure 14. A continuum of the purpose of using technology and its users within the TPACK rubric.

As one moves towards the middle level (adapting) a shift in how the teacher is using technology will be noticed. Here, teachers begin to see the cognitive value of technology. They are not just using technology for the sake of using it; rather, they are trying to use it in a way that helps students deepen their understanding of the material. That is, technology is being used more as a cognitive tool, and there is a new and different approach to teaching and learning that would not otherwise be possible without technology. At this level, exploration and inquiry begins, though in a limited sense, as the teacher is still the primary user of technology. For example, a teacher may use a web applet to showcase the construction of a sampling distribution through the use of a

simulation. This dynamic piece of constructing the simulation is the "new and different" approach (as opposed to merely displaying the end result of a simulation).

Once teachers begin allowing their students to be the primary users of cognitive technologies, they begin to progress to the two highest levels in the TPACK rubric (exploring and advancing). At these levels, *students* are the primary users of technology and use it as a cognitive tool to explore and help develop their conceptual understanding. For instance, rather than displaying a randomization distribution, a teacher can have students define their own probability model and create the distribution using a simulation – at the same time asking thought-provoking questions that require students to reflect upon what they are learning and doing with technology. At these levels, a teacher serves as a facilitator, rather than a director, who focuses on students' conceptual understanding of the material.

One of the first changes I made to the original TPACK rubric was combining the two highest levels. It is clear that teachers at either of these "top" levels have reached a point where they have a strong understanding of the value and purpose of integrating technology and inquiry into the classroom. This is immediately apparent in the language of the original authors (Tournaki & Lyublinskaya, 2014), as shown in Table 3.

I have bolded the slight changes in language between the two descriptions. I noticed that "aligned" is replaced by "strongly aligned," "important" by "essential," and "appropriate" by "effective." Clearly, these distinctions can be exceedingly difficult to recognize in a classroom observation. Moreover, they are essentially a description of (for the goals described herein) the same thing; that is, that technology is fully aligned with

the goals of the curriculum, and a teacher has chosen topics that can be effectively taught with technology. For this study, I made no effort to distinguish between these two top levels. Ultimately, at these levels, a teacher has demonstrated a higher understanding of how to effectively integrate technology in the classroom. However, I can distinguish between the three lowest levels (recognizing, accepting, and adapting).

Table 3. Comparing the language between the top two levels (exploring and advancing) of the original TPACK rubric within the curriculum component (Tournaki & Lyublinskaya, 2014).

Exploring (4)	**Advancing (5)**
Technology is **aligned** with curriculum goals. Teacher chooses **important** topics of school mathematics or science curricula and technology use is **appropriate** for the chosen curriculum topics.	Technology is **strongly aligned** with curriculum goals. Teacher chooses **essential** topics of school mathematics or science curricula. Technology use is **effective** for the chosen curriculum topics.

Next, I chose to retain the two performance indicators in each cell. Instead of using language that is predominantly subjective, I added concrete descriptors and examples in each performance indicator. This was essential to ensure consistency when using the rubric, as other researchers were testing the rubric with me.

Finally, I revised the possible "technology score" in each of the four components (rows) to be a *whole* number from 0 to 4, where 0 denotes not meeting the standards of the lowest level. Rather than receiving a half-point for meeting each performance indicator, I considered both performance indicators together; I recorded one point if a combination of the performance indicators had been met. I recorded a "technology score" for each of the four components, and defined the overall TPACK level to be the sum of scores across all four components.

The stated purpose of the TPACK Levels Rubric (Lyublinskaya & Tournaki, 2012), is to "assess teachers' TPACK levels based on qualitative data collected from teachers, such as written artifacts, descriptions of observed behaviors, and teachers' self-reflections." Therefore, in this study, I used an adapted TPACK rubric to evaluate the qualitative descriptions of classroom observations.

Implementation of the Statistics Lessons

Within each statistics unit are five lessons, and I asked participants to implement four of these lessons (Lessons 2 – 5) during the 2015-2016 academic school year. Six of the eleven participants were unable to teach the statistics unit by the end of the 2015-2016 academic school year. Consequently, they agreed (in June 2016) to teach the statistics lessons at the beginning of the 2016-2017 academic year. These four lessons need not be taught continuously, but I asked participants to teach the lessons in order (Figure 15). For each lesson, classroom observations were employed and I recorded five technology scores (one for each component and a sum) for each participant for each lesson.

The only difference between the annotated and non-annotated statistics units can be found in Lessons 3 and 4. While these lessons cover the same content in both units, those in the annotated statistics unit contain annotations. These are also the only two lessons (in both units) to contain technology-based activities. I used Lesson 2 to assess how all participants teach statistics with technology and inquiry *before* implementing lessons with annotations (if applicable) and technology-based activities. Lessons 3 and 4 were meant to help guide teachers who received the annotations to focus on developing

students' conceptual understanding through the effective use of technology and, in particular, simulation. Finally, I used Lesson 5 to evaluate how participants teach statistics with technology *after* implementing lessons with annotations (if applicable) and technology-based activities.

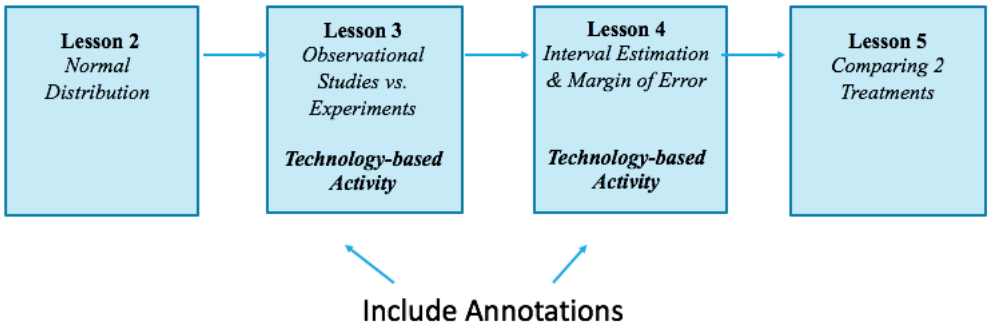


Figure 15. The order in which the statistics lessons were implemented and the content of each lesson.

Post-Implementation of the Statistics Lessons

After teachers implemented the four designated lessons, I asked them to complete two final surveys. I used the first survey to assess their current thinking (post-experiment) about the use of inquiry and technology in their classroom and the use of simulation to teach statistical concepts. I used the second survey to obtain background information about their educational and professional experience. In addition, I asked selected participants to complete an exit interview.

Data Collection

For the purpose of collecting various forms of data to help investigate the three research questions, I used both quantitative and qualitative measures. The research

instruments included multiple surveys (given before and after participants implemented the four statistics lessons), classroom observations, and interviews.

Surveys

Four surveys were given to participants throughout this study (Appendix D). I used the surveys to gather background information to recruit and sort participants into blocks; they were also used to gain insight into teachers' statistical content knowledge and use of technology and inquiry in a mathematics classroom. The enrollment and initial surveys were given pre-experiment (before lessons were implemented), and the exit and teacher background and experience surveys were given post-experiment (after lessons were implemented). I administered all of these surveys online and gave participants a unique 3-digit code to use as an identifying "user name" in the surveys.

Enrollment Survey. To recruit participants, I created a brief enrollment survey calling for participation in my study. In this survey, I asked teachers to provide their names, work e-mail addresses and home states, and to identify whether or not they would be teaching Algebra II, Mathematics III, or the equivalent course during the 2015-2016 academic school year.

Initial Survey. I administered the initial survey to all interested participants (teaching the appropriate course) prior to the start of the experiment and immediately after they completed the enrollment survey. It consisted of four parts: (1) a statement of informed consent; (2) information about the individual's district, school, class schedule, and accessibility to class sets of computers or tablets; (3) current awareness about the use

of technology and inquiry in the individual's mathematics classroom; and (4) a statistical content knowledge assessment. The primary purpose of this survey was two-fold: to determine the final eligibility of participants and collect information to sort them into as similar of groups as possible for treatment allocation.

The statement of informed consent explained the context, research plan and timeline, risks and benefits of the research study, and confidentiality. Participation in this study was voluntary, and the consent form acknowledged this and stated that participants may choose to relinquish their role in the study at any time.

To establish the final eligibility of participants, I asked interested participants about their access to class sets of computers or tablets. I also prompted participants to disclose their 2015-2016 teaching schedule and the names of their district and school. I asked the latter questions to ensure only one teacher per district was selected to avoid potential "contamination" of the treatment group. It is often the case that teachers talk with one another and discuss their classes; contamination occurs when participants assigned to a control group discuss the study with those assigned to a treatment group, inadvertently absorbing some of the treatment as a result. Hence, I decided to allow only one participant per district. In the event there was more than one teacher per district willing to participate in the study, I would randomly select one to be a participant. However, this did not occur.

Next, I asked six open-ended questions prompting potential study participants to express their views (pre-experiment) about teaching and learning mathematics with technology and inquiry explorations; two of these questions focused on the use of

inquiry, while the rest focused on technology. For example, one question asked participants to describe *how* they have used technology in their classroom, focusing on the most *productive* use of it. By emphasizing a productive use of technology, I was able to glimpse into how participants viewed the use of technology – that is, whether they saw it as primarily a cognitive tool or a productivity tool. I used the responses to these open-ended questions, in part, to sort participants into blocks. These responses also provided information about participants' views pre-experiment, serving as a basis to compare their views post-experiment.

I created the open-ended questions using an iterative process. The questions were first adapted from those Mudzimiri (2012) used in her TPACK Survey to gather information about pre-service teachers' understanding of the components of the TPACK framework. Next, the questions were discussed and altered with an outside researcher knowledgeable about the TPACK framework. Once I developed the questions, they were sent to local high school teachers and graduate students (many of whom were currently in-service high school teachers) enrolled in a mathematics education online graduate course. The purpose of this was to investigate whether or not the questions posed were prompting answers that aligned with the technology and inquiry-related goals of this research. I was curious, for example, whether teachers would respond to these questions in ways that do not reflect *their* implementation in their own classroom. For instance, an original question on this survey was

> *"Describe and compare a good use of technology in teaching mathematics to a trivial or unproductive use of technology in teaching mathematics."*

I revised this question to

> *"If you have used technology in your mathematics classroom in the past 6 months, describe how you have used it. Focus on what you believe to be the most productive use of technology. If you have not used technology in teaching mathematics, please explain why."*

Based on the initial responses from this group of teachers, I revised the questions a final time.

Finally, I administered a statistical content knowledge assessment because I expected participants to have varying degrees of statistical content knowledge. To assess this knowledge, I used resources from the National Science Foundation-funded ARTIST project (Garfield et al., 2006). The ARTIST project focuses on the conceptual understanding of statistics along with assessing statistical literacy, reasoning, and thinking; since these goals align with those presented in the *Pre-K-12 GAISE* and *SET Reports*, it seemed natural to use this resource. The ARTIST project provides multiple, free online assessments. There are 11 multiple-choice tests on individual topics that include data collection, data representation, measures of spread, measures of center, normal distribution, probability, bivariate quantitative data, bivariate categorical data, sampling distributions, confidence intervals, and significance tests. Each test consists of 7 to 15 items and takes approximately 10 to 15 minutes to complete.

In introductory statistics, sampling and randomization distributions are generally considered central to making inferences using hypothesis testing and calculating confidence intervals. Because the statistics unit I created for this study focuses on

statistical inference and making conclusions, I wanted to assess participants' knowledge about sampling and randomization distributions. Therefore, I chose to administer the ARTIST assessment regarding sampling distributions to participants; this particular assessment contains 15 multiple-choice questions and was scored based on correctness. I recorded a point for each correct answer, and used the total number of problems answered correctly (out of 15) as a measure of a teacher's statistical content knowledge.

Exit Survey. After participants implemented the four statistics lessons, I asked them to complete an exit survey. This survey consisted of the same six open-ended questions from the initial survey regarding the use of technology and inquiry and one additional question regarding their views on using simulation to teach statistical concepts. I used this survey to assess each participant's post-experiment views about the use of technology and inquiry in a mathematics classroom, and I compared these with the participants' pre-experiment responses. Furthermore, because technology is a tool often used to perform a simulation, the participants' views about the use of simulation in teaching statistics may provide further insight to their views about using technology in a mathematics classroom. I used the responses to the exit survey to help build and describe a story that depicted any changes in the participants' views and awareness of the use of technology and inquiry in their mathematics classrooms.

Teacher Background and Experience Survey. I gave the teacher background and experience survey post-experiment. It consisted of questions about the participant's educational background, particularly in relation to statistics. In addition, I asked for

demographic information from their school, along with brief questions about their professional experience and whether or not they had undergone any professional development in the use of technology or teaching statistics. I used this information to provide background information about the participants, aiding in truly understanding the common and disparate characteristics of the final group of participants.

Classroom Observations

Observations allow a researcher to grasp how individuals piece together aspects of their lives (Savin-Baden & Major, 2013; Creswell, 2013). Through the use of classroom observations, I was able to document *how* participants used technology and inquiry in their classrooms and the explicit decisions they made when using them.

Multiple observations of the same teacher are required to obtain a more complete picture of how teachers synthesize their knowledge about technology and inquiry into their classrooms. Accordingly, I observed each participant teach four statistics lessons, each ranging from one to two days in length. With the exception of one teacher, these lessons were taught continuously during a one- to two-week period. A few participants taught several sections of Algebra II or Mathematics III. In these cases, one of their class periods was randomly selected to use for classroom observations.

Due to participants being broadly spread across the U.S., observations occurred via video-recordings or FaceTime. I provided those recording videos of themselves teaching with instructions on what to record, how to record, and when to record. I requested each participant record the entirety of each lesson, placing the camera in a position where the whiteboard and front of the classroom could be viewed with ease. The

participants had the option of placing the camera in the back of the classroom or the front of the classroom (to avoid students being filmed). The participants transferred any video files via a secure Dropbox or Google Drive folder; these files were kept on a personal computer only I had access to.

Field notes and an observation protocol supplemented the classroom observations and helped me record "technology scores" for each participant and lesson. I took field notes during each observation and they mostly revolved around implementation of the lesson. They included things such as the type of technology being used and how the teacher and students were using it, how inquiry was being implemented in the classroom, the statistics content being taught, how the participant was or was not using the annotations, how the lesson plans were followed, any struggles the participants or their students encountered during the lesson, the types of conversations and interactions occurring, and generally any teaching strategies or student misconceptions and questions occurring throughout the lesson.

Following the classroom observations and field notes, I completed an observation protocol (Appendix E). Here, I summarized the field notes into the following categories: implementation, struggles with content, struggles with technology, interaction with students, and other, aiding in organizing the observation data. More importantly, all of these notes helped me evaluate each teacher's use of technology as well as identify possible avenues of revision of the annotations and lesson plans to pursue in the future.

Interviews

Following the implementation of the statistics lessons, I conducted semi-structured, primarily audio-recorded interviews, lasting approximately 20 to 60 minutes. In semi-structured interviews, pre-established questions are asked, though one may ask additional questions (tailored to the interviewee's responses), affording some necessary flexibility in interview structure and time use (Savin-Baden & Major, 2013).

The primary purpose of these interviews was to gain additional insight in lesson implementation and the participant's perspectives on the lessons (including the annotations and activities), teaching statistics, and using technology and inquiry. In some cases, the classroom observation videos were incomplete, damaged, or non-existent; here, the interviews were able to provide key insight. Generally, these interviews revealed a participant's logic behind established class norms and structure, process of lesson implementation, and instructional strategies.

In total, I conducted three sets of interviews. Not all participants were selected for an interview. Those with incomplete videos were preferentially chosen so I could "fill in the gaps." I also contacted others based on their use of the annotations and implementation of the lessons. For instance, I observed one participant using the annotations, so I contacted that participant to ask about the usefulness of the annotations. I also contacted another participant who did not use the annotations to ask similar questions.

Summary of Data Collection

Overall, I collected data before, during, and after the experiment, ranging from May 2015 to September 2016 (Figure 16). I also briefly summarize the purpose of each source of data I collected (Table 4).

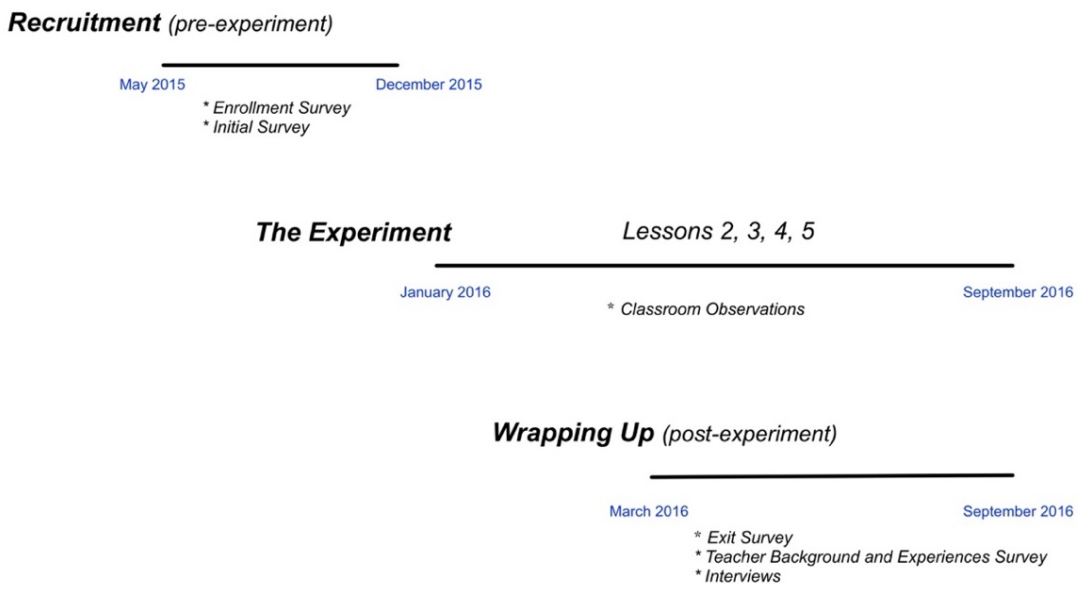


Figure 16. A timeline of when I collected data.

Table 4. A summary of the research instruments I used to collect data.

Data Source	**Purpose**
Enrollment Survey	• Find out if interested participant will be teaching Algebra II, Mathematics III, or the equivalent during the 2015-2016 academic school year.
Initial Survey	• Verify interested participant have access to a class set of Internet-enabled computers or tablets. • Sort final group of participants into blocks. • Gather information about a participant's awareness of the use of technology and inquiry *pre*-experiment.
Teacher Background and Experience Survey	• Gather information about a participant's school and educational and professional background.
Exit Survey	• Gather information about a participant's awareness of the use of technology and inquiry *post*-experiment. • Gather information about a participant's view regarding the use of simulation to teach and learn statistical concepts.
Classroom Observations	• View the implementation of the four statistics lessons within each unit. • Evaluate how each participant used technology and inquiry during each lesson. • Provide suggestions for revisions to annotations and lesson plans.
Interviews	• Acquire additional insight in how the lesson plans and annotations were implemented. • Gather information about the usefulness of the annotations.

Analyzing the Data

I used both quantitative and qualitative methods of analyses to analyze the data before, during, and after the experiment. Different sources of data were used to recruit, select, and sort participants into blocks and address the three research questions stated at the beginning of this chapter (Table 5).

Table 5. The different sources of data I used to recruit, select, and sort participants and address each research question.

		Data Sources					
		Enrollment Survey	Initial Survey	Exit Survey	Teacher Background and Experience Survey	Classroom Observations	Interviews
Experimental Design	• Recruit and select participants. • Sort participants into blocks.	✔	✔				
Research Questions	1. To what extent do specially annotated lesson plans enable in-service high school teachers to effectively integrate technology in their instruction of statistical concepts?					✔	✔
	2. Do the qualities inherent in specially annotated lesson plans carry over into in-service high school teachers' implementation of a future technology-assisted statistics lesson?					✔	✔
	3. How are the following qualities associated with an in-service high school teacher's use of specially annotated lesson plans as a means to guide the instruction of statistics with technology?		✔	✔	✔	✔	✔

Pre-Experiment Data Analysis

The data I collected from the enrollment and initial surveys prior to the experiment were primarily used to select participants and sort them into homogeneous

groups. From the initial survey, I recorded a statistical content knowledge score for each participant. I then used the participants' responses to the six open-ended questions about technology and inquiry to first categorize and then rank participants on their awareness of the use of technology and their awareness of the use inquiry in a mathematics classroom. As part of the development of the experimental design, I sought to establish how the treatment conditions would be allocated prior to the start of the experiment. I began by involving all 28 interested participants to first establish the method I would use, and then I applied this method to the 11 participants who received a statistics unit.

I individually examined the responses (on technology, inquiry, and statistical content knowledge) from the original 28 teachers who expressed interest in participating and had met the stated criteria of the study. First, I synthesized their full responses to the two inquiry-related questions and then condensed their responses with primary consideration of their awareness of the use of inquiry to teach and learn mathematics. Upon compiling each participant's summarized responses in an individual table, I sorted these responses into two groups: "high inquiry," indicating a more nuanced awareness of using inquiry in the classroom, and "low inquiry," indicating less awareness. I used a similar method to condense the four open-ended technology-related questions. Here, I established what teachers viewed as productive uses of technology and the role of students and the teacher when using technology. Again, I compiled all of the participants' summarized responses in a single table and sorted them into two groups – "high technology" and "low technology," indicating the broad levels of awareness teachers conveyed of using technology to effectively teach mathematics. Next, I created a plot

representing how participants were sorted into the four "high/low" combinations of technology and inquiry (Figure 17). I created four side-by-side boxplots, with each boxplot representing one of the four combinations. The individual statistical content knowledge scores overlay the boxplots and are represented by the participant's unique character identifier.

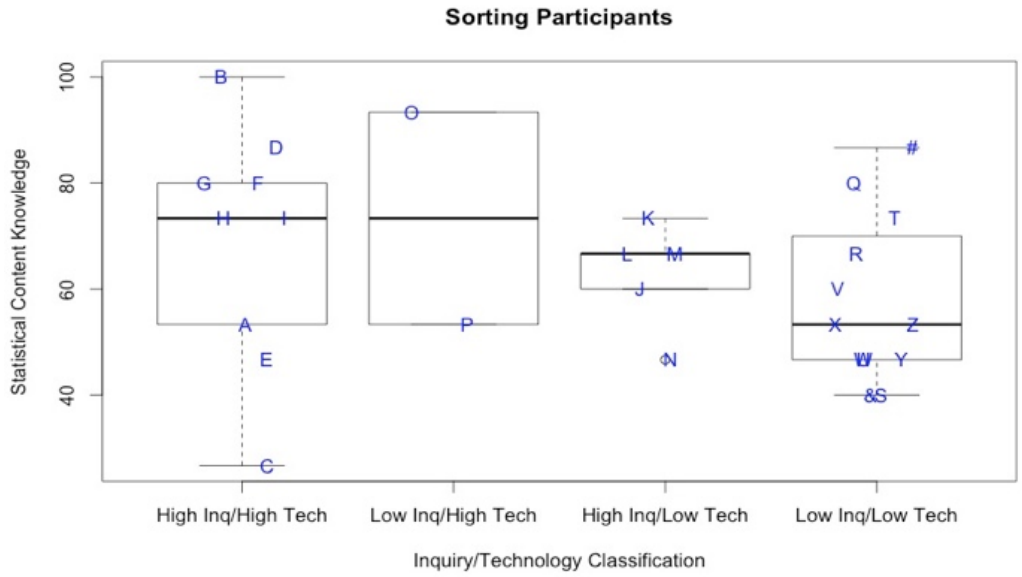


Figure 17. Sorting participants into four blocks by their awareness of the use of technology and their awareness of the use of inquiry. The participant's statistical content knowledge scores overlay the boxplots and are represented by a unique character identifier.

I initially used this particular strategy to sort participants into blocks. At this point, participants within each block would then be randomly assigned a treatment condition. However, I revised this strategy to treat the three teacher characteristics (statistical content knowledge, awareness of the use of technology, and awareness of the use of inquiry) more continuously.

The new (and final) strategy still included the participants' statistical content knowledge scores, but participants were now *ranked* on their awareness of the use of technology and their awareness of the use inquiry in a mathematics classroom. To do this, I examined the tables of condensed responses (described above) as a whole, ranking responses within the table from 1 ("high") to 28 ("low"); I also discussed and verified these rankings with an outside researcher familiar with the research study.

For example, consider the following summarized responses from teachers about the use of inquiry for teaching and learning mathematics.

> **Teacher 1:**
> "Exploration and inquiry are very important in my advanced classes not so much in geometry. I haven't used it [inquiry] much because in standard classes there is a lot of cumbersome work and very little gain in understanding by students."
>
> **Teacher 2:**
> "Have never used them [exploration and inquiry]."
>
> **Teacher 3:**
> "Occasionally when I need to introduce a difficult topic, I resort to exploration learning so students can acquire knowledge themselves. Under such a constructivist model, students forge connections in the subject matter on their own. From personal experience, I have noticed that students are more willing to retain knowledge that they actively discover; rather than merely parrot."

In these responses, there is a clear ordering of teachers' awareness of the use of inquiry and exploration in a mathematics classroom. Teacher 2 demonstrates the least amount of awareness of the use of inquiry. Teachers 1 and 3 indicate a higher level of understanding; however, Teacher 3 reveals a more nuanced understanding than Teacher 1. While both of these teachers appear to use inquiry in their classrooms, they have a distinct interpretation of when to use it. Teacher 1 states how inquiry is beneficial in more advanced classes as opposed to standard classes (such as Algebra II). Teacher 3 also

acknowledges a benefit of using inquiry in the classroom stating how students are acquiring knowledge themselves rather than simply being told.

After each participant's technology and inquiry responses were ranked in this fashion, I compared the technology and inquiry rankings to the four original blocks (Figure 17) to check for consistency. I found those ranked higher in inquiry and technology were in fact, in the high inquiry/high technology block. Similar results were found for the other blocks, confirming the rankings were reasonable and consistent with previous blockings.

As I was now treating the three teacher characteristics (technology, inquiry, and statistical content knowledge) as continuous variables, I used a different type of plot to display the rankings and statistical content knowledge scores. A bubble plot is one way to display three dimensions of data; it extends a scatterplot to include a third dimension of data represented by the size of the plotting symbol (i.e., bubbles) where the size of the symbol is proportional to the value. I used this plot to simultaneously display the statistical content knowledge scores, technology rankings, and inquiry rankings (Figure 18). In this figure, participants are represented by the characters "A-Z", "#" and "&." The x-axis represents the *technology* ranking, the y-axis represents the *inquiry* ranking, and the size of the bubble surrounding each character is proportional to the participant's statistical content knowledge score. The higher the statistical content knowledge score, the larger the size of the bubble. The rankings were from 1 (high) to 28 (low).

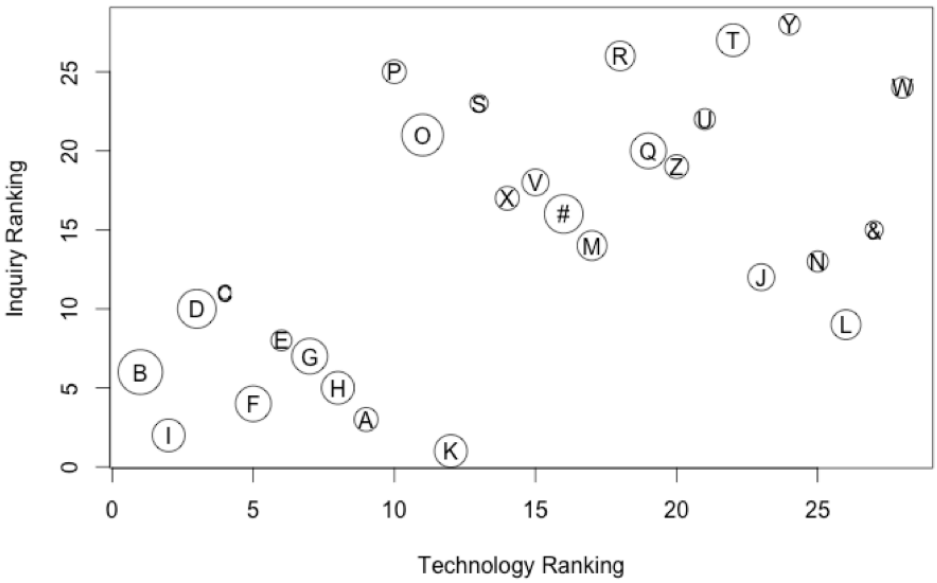


Figure 18. Statistical content knowledge scores, technology rankings, and inquiry rankings for the 28 participants initially interested in the study. Participants were ranked 1 (high) to 28 (low) on their awareness of the use of technology and their awareness of the use of inquiry. The size of the bubble surrounding a character identifier is proportional to the participant's statistical content knowledge score.

I used the information displayed in Figure 18 to sort participants into as similar of pairs as possible (Figure 19), selecting pairs based on similarity of the three teacher characteristics of technology, inquiry, and statistical content knowledge. All the pairings in Figure 19 were discussed with other researchers. The purpose of this process was to corroborate the pairings and ensure consistent groupings among various researchers.

There are many ways one could create pairings from Figure 18 and there is no single correct pairing. Rather, the pairings reflect different beliefs regarding the three characteristics of technology, inquiry, and statistical content knowledge. In fact, the bubble plot could have been created differently by changing the variables on the axes (e.g., Figure 20). Placing the three characteristics on different axes may lead one to choose different pairings.

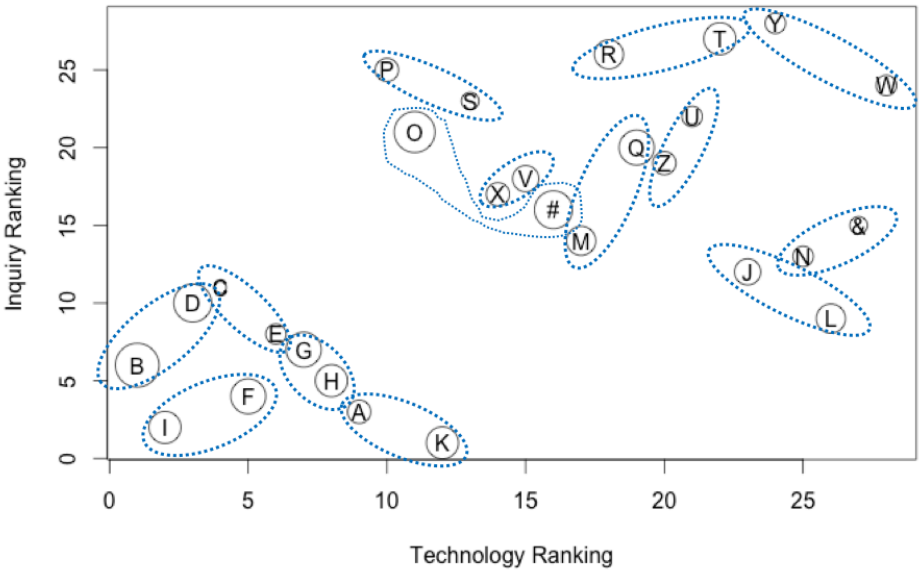


Figure 19. The 28 participants sorted into pairs with similar rankings for technology and inquiry, as well as similar statistical content knowledge scores.

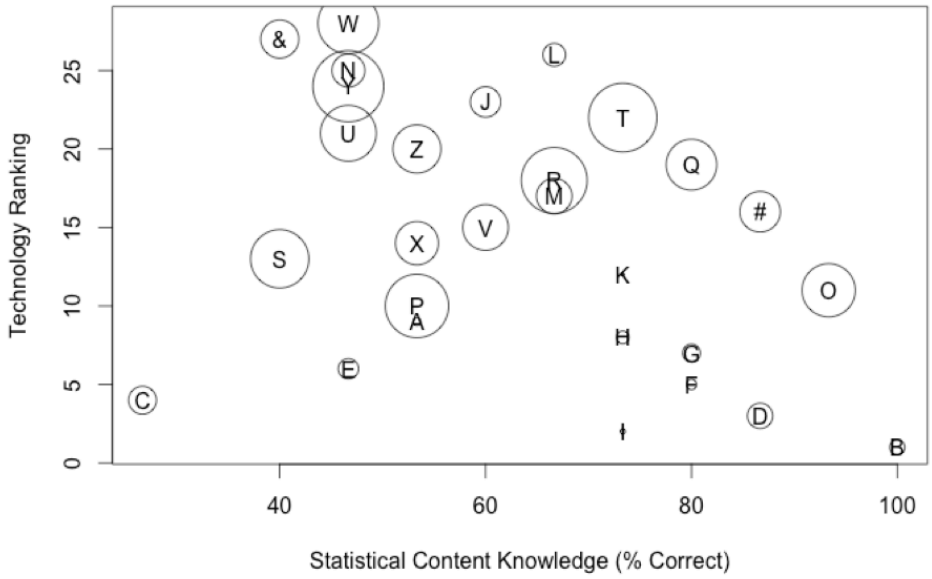


Figure 20. A different way to create the bubble plot. Here, statistical content knowledge is the x-axis, technology ranking is the y-axis, and the size of the bubble corresponds to the inquiry ranking; smaller bubbles correspond to higher inquiry rankings.

However, I purposely choose to construct the bubble plot by placing the technology and inquiry rankings on the x- and y-axis, respectively, and representing the statistical content knowledge scores as bubbles (Figure 18). The TPACK rubric I used to evaluate classroom observations is centered on the use of technology and inquiry. Consequently, those who use technology and inquiry more effectively will generally obtain higher scores. While statistical content knowledge is necessary to teach statistics, it is not directly reflected in the TPACK rubric. Therefore, with regard to using the TPACK rubric, I viewed the technology and inquiry rankings as more important, and I wanted to consider statistical content knowledge *after* I first considered the technology and inquiry rankings. To do this, I chose to display the technology and inquiry rankings on the primary axes, allowing me to more easily consider their proximity. In addition, the rankings were both from 1-28. Therefore, the x- and y-axis were scaled equivalently, making for easier comparisons.

By constructing the bubble plot a different way (e.g., Figure 20), I would have been more inclined to first compare the technology ranking and statistical content knowledge score before considering the size of the bubbles. In this construction, however, the scales differ on the x and y-axes; the statistical content knowledge scores range from 27% to 100% and the technology rankings are from 1 to 28. A 1-unit change in the y-axis (technology ranking) is different from a 1-unit change in the x-axis (statistical content knowledge) making it more difficult to compare these two characteristics.

I applied this ranking and pairing process to the final 11 participants who agreed to participate (Figure 21). Because of the odd number of participants, I created one triplet (e.g., Participants "H," "G," and "E") and paired the remainder of the participants.

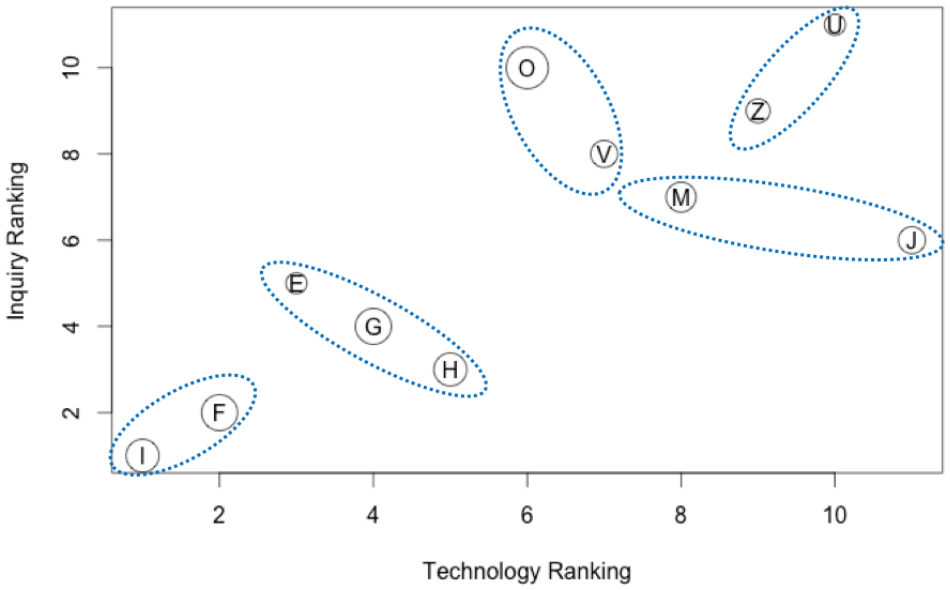


Figure 21. Ranking and sorting the final 11 participants into similar blocks based on technology, inquiry, and statistical content knowledge.

Post-Experiment Data Analysis

I performed the more formal analysis of the research study after participants implemented the lessons and I used both quantitative and qualitative methods to analyze the data. Classroom observations were the primary data collected in this study, and I quantitatively analyzed participants' technology scores to investigate the first two research questions and lend insight to the third research question. To supplement this analysis, I also employed a qualitative approach; together with interviews and surveys, select classroom observations were analyzed in this way.

Classroom Observations. As I observed Lessons 2 through 4, I took field notes on each day of the lesson. After each lesson was implemented in its entirety, I reflected on the entire lesson and rewrote the field notes, making a point to include additional observations as necessary. From here, I reflected on the lesson and summarized my notes in an observation protocol. The combination of these two (field notes and observation protocol) provided me with a vivid picture of how each lesson was implemented. However, in the absence of complete observations, I used the information from interviews to fill in the observation protocol. In the end, I evaluated each of the four lessons with the adapted TPACK rubric. For each lesson, I recorded five technology scores per participant – one for each of the four components of the TPACK rubric (C1: *Conception*, C2: *Students*, C3: *Curriculum*, and C4: *Strategies*) and a total of these four score (Table 6).

Table 6. An example of technology scores recorded from the classroom observations.

Participant	**Lesson**	**C1**	**C2**	**C3**	**C4**	**Total**
001	2	2	4	3	1	10
002	2	1	1	2	0	4
003	2	4	4	3	4	15

To supplement the numeric technology scores, I wrote a brief justification for each score. After all evaluations of a single lesson were completed, I compiled all of the participants' scores and justifications in a single table and re-examined the scores as a whole to check for consistency *within* lessons. For instance, if two participants demonstrated how to use graphing calculators to calculate normal percentiles and gave their students problems to practice this technique on, then these two participants should

have received the same technology score in the appropriate component. A similar strategy was used to check for consistency *between* lessons.

The goal of the first research question: "To what extent do specially annotated lesson plans enable in-service high school teachers to effectively integrate technology in their instruction of statistical concepts?" was to investigate what happens, in a given lesson, when a teacher is provided with an annotated lesson plan. That is, is it possible to capture efficiency of implementation so that other teachers with access to these annotated lesson plans can demonstrate a heightened ability to effectively integrate technology while teaching statistics? The annotated lesson plans (found in Lessons 3 and 4 of the annotated statistics unit) and corresponding technology-based activities were designed with a particular goal in mind; if the lessons were properly implemented, participants would be effectively integrating technology and receive a technology score of "4" in each of the four components, constituting a total technology score of 16.

The technology scores from Lesson 4 were specifically used to investigate the first research question. To do this, I compared the technology scores from those who received annotations to those who did not. Even though both Lessons 3 and 4 contained annotations (in the annotated unit) and a technology-based activity, I wanted to use the technology scores from Lesson 4 in the analysis, as these scores would better represent how teachers used the annotations and/or technology-based activity. Teachers would have had practice using the web applet and annotations (if applicable), minimizing the amount of difficulties they may experience implementing a new type of technology.

The second research question: "Do the qualities inherent in specially annotated lesson plans carry over into in-service high school teachers' implementation of a future technology-assisted statistics lesson?" focused on the change in a participant's technology scores from Lessons 2 to 5. One of the goals of this research question was to investigate whether annotated lesson plans are a means to develop a teacher's effective use of technology when teaching statistics. In other words, do teachers learn ways to effectively integrate technology from the annotations, and do they implement this knowledge in future instruction of statistics (where technology and inquiry can aid in the development of students' conceptual understanding)? As Lesson 2 occurred before and Lesson 5 occurred after participants implemented annotated lesson plans (if applicable) and technology-based activities, I compared the difference in technology scores from the second to fifth lesson between treatment groups to explore the possibility of teachers applying knowledge from the annotations to a future statistics lesson.

Exploratory Plots of Classroom Observations. To begin with, I created plots to visualize the participants' five technology scores from each lesson and to identify any patterns in the data. Because I recorded multiple technology scores for each participant, I obtained several response measurements for each participant (multivariate response). One type of plot particularly useful for visualizing and graphically exploring multivariate data is a profile plot. It allows one to visualize and compare the technology scores within and between participants and identify any patterns. Profile plots are also helpful for gathering information about the structure of a model, providing insight for how data are to be analyzed. I used profile plots to investigate how to use the technology scores in my

quantitative analysis. These plots helped me decide whether to analyze each of the four technology scores (one per component) separately or use the total of those scores.

A profile plot is a particular style of a point and line plot that is drawn separately for each participant, displaying, for instance, the technology scores versus the four components of the TPACK rubric (Ramsey & Schafer, 2013). Each data point from a *single* participant is connected with a line, meaning the lines are connecting technology scores from the *same* participant. This type of plot allows one to clearly view the technology scores for each individual participant. With a small number of participants, this is an effective way to visualize how participants vary in their technology scores and investigate patterns between and within participants, ultimately guiding the quantitative analysis.

For example, I used hypothetical technology scores (for Lesson 2) to create two possible profile plots (Figures 22 and 23). I have two vertical panels in each profile plot – one solely displaying the technology scores from participants who received the annotated statistics unit and the other displaying the technology scores from those who received the non-annotated statistics unit. In these plots, the four components are displayed on the x-axis and the corresponding technology scores are on displayed on the y-axis. Each line represents a single participant and contains four points corresponding to the four technology scores (from the four components) for that participant. Participants who were paired together have the same colored line.

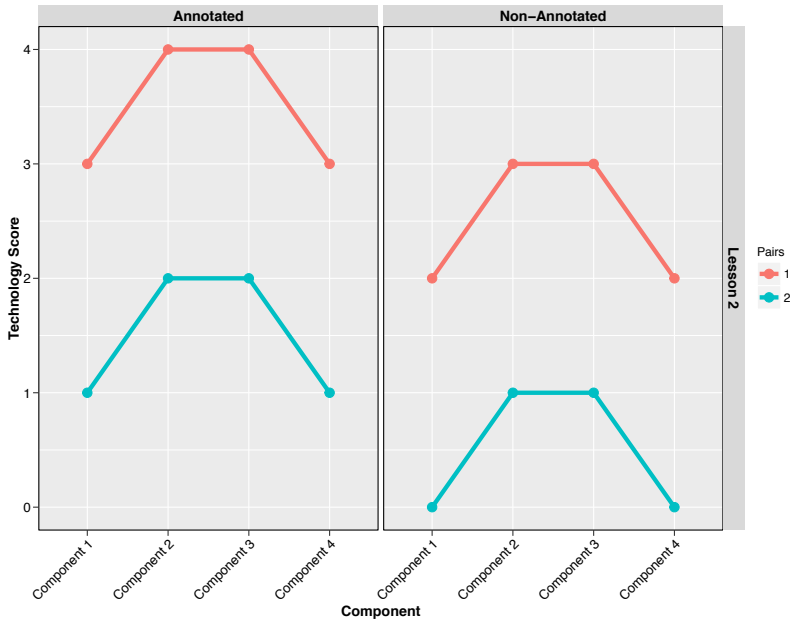


Figure 22. An example of a profile plot (using hypothetical data) with *parallel* lines.

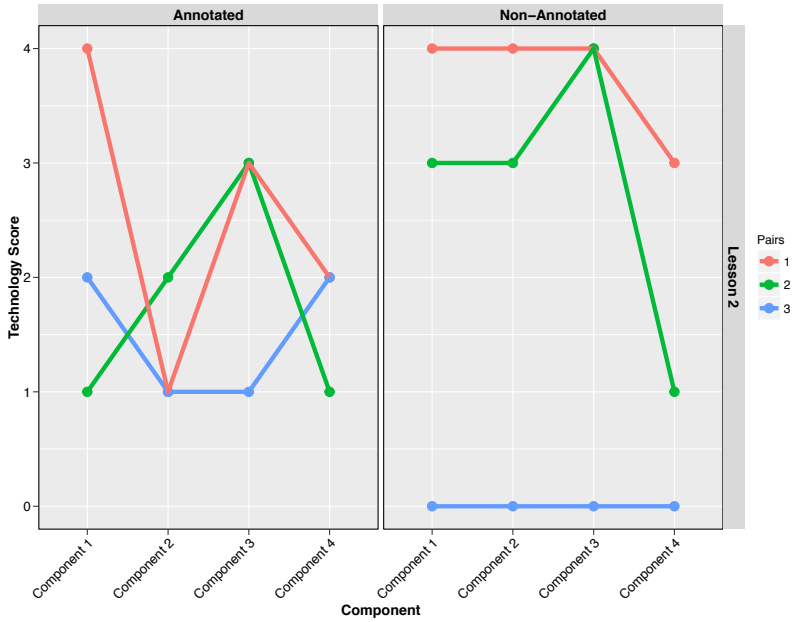


Figure 23. An example of a profile plot (using hypothetical data) with *non-parallel* lines.

These plots were, in part, used to investigate the dependence among the technology scores for the four components from an individual (per lesson) and how the participants' technology scores for each component (per lesson) fluctuated together. That

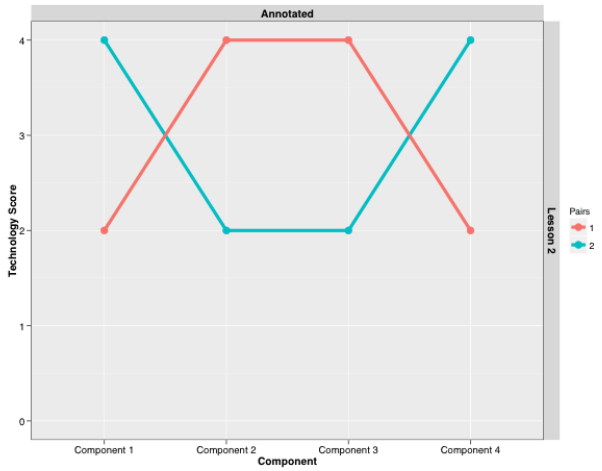
is, are participants' technology scores on a component consistently related to the score on another component? I used these types of plots to help choose between separate analyses on the technology scores from the four components or an analysis on the total (or some other aggregate measure, such as an average) of the four technology scores per lesson. It was my initial belief that annotations would have a different effect on individual components of the TPACK rubric, thereby affecting the technology scores recorded for each component. Therefore, I initially wanted to use the four technology scores separately, rather than using an aggregation of these scores. However, if the profile plots suggested the technology scores were dependent on the four individual components, I would be adding more complexity to the analysis by analyzing the technology scores separately.

A profile plot with parallel lines (Figure 22) suggests dependence among the technology scores for the four components. In other words, the relationship between the technology scores and each component (in a lesson) is the same for all participants. Therefore, a single aggregate score (e.g., the total of the technology scores from the four components of a lesson) would be an appropriate response to analyze because the separate technology scores from each component do not change the story about a participant's use of technology and inquiry. Here, the technology scores for each component behave similarly for each participant, thereby making the total of the technology scores a quantity that serves as a fair standard of comparison. For instance, a profile plot with approximately parallel lines indicates a participant with a higher total technology score also had higher technology scores recorded for each individual

component. In this situation, I could choose to account for this dependency by including a correlation structure in a statistical model (adding complexity), or I could choose to simplify the model by using the sum of the four technology scores as the response variable for each individual, ultimately ignoring the individual components.

Conversely, a profile plot with non-parallel lines (Figure 23) suggests lack of dependence among the technology scores for the four components. That is, the relationship between the technology scores and each component (in a lesson) is different for all participants. Therefore, within an individual, one does not gain the full story on the use of technology based on a total of the four component scores. For example, consider the following (hypothetical) technology scores (for Lesson 3) for two participants who received the annotated statistics unit (Figure 24).

Both participants have a total technology score of 12, indicating the same "level" of using technology, yet their four component technology scores exhibit different behavior. Their four component technology scores, depicting how they used technology in the classroom, are not fully represented by a total. Rather, the component technology scores should be analyzed separately (or allowed to vary separately in a more complex model including correlation) to describe the complete story of how teachers are integrating technology into the lessons. In this situation, I balance model complexity and appropriateness to analyze each of the four component technology scores separately to capture the possibly different stories for the components.



Participant	C1	C2	C3	C4	Total
001	2	4	4	2	12
002	4	2	2	4	12

Figure 24. An example of two participants' (hypothetical) technology scores. Their total technology score is the same, yet they display different scores in each component.

Statistical Analysis of Classroom Observations. Only five of the 11 participants completed the study and implemented the statistics lessons. Therefore, due to the very small number of participants who completed the study, a formal statistical analysis of the classroom observations was not feasible to employ in investigating the research questions. The remainder of this section details the type of statistical analysis I would have performed had I retained all 11 participants.

The technology scores from the classroom observations were from a completely randomized block experiment, in which participants were sorted into pairs and a triplet (i.e., blocks). Then, within each block, participants were randomly assigned to receive the annotated or non-annotated statistics unit. To investigate the first research question (examining what occurs, in a given lesson, when a teacher is provided with an annotated lesson plan), I planned to use technology scores from Lesson 4 as my response variable in the statistical analysis. For the second, focusing on applying knowledge from the

annotations to a future statistics lesson, I was going to use the difference in technology scores between Lesson 2 and Lesson 5 as my response.

Given the nature of the design of this research study, using a randomization distribution to make inferences would have been a natural method for analysis. Furthermore, this method of inference does not rely on any normality assumptions, unlike an ordinary linear regression analysis. Because the response variables (technology scores) are integers and there are few participants in the study, these scores are quite nonnormal, clearly violating the normality assumption of a t-distribution approximation (i.e., regression analysis). Given the categorical nature of the response variables, an ordered categorical regression would also be an alternative method to analyze these data (Agresti, 2007).

In the context of this study, a randomization distribution is a distribution of the response of interest after considering all possible re-randomizations of the treatment assignments within the blocks. For example, suppose (in one pair) *Participant A* was randomly assigned to receive the annotated statistics unit and *Participant B* received the non-annotated unit. There are two possible ways ($2\ choose\ 1 = 2$) to randomize the treatment conditions within this pair (Table 7). By chance, the original treatment conditions could be re-assigned to the two participants or the opposite treatment conditions could be re-assigned. This process would need be completed for *each* pair of participants in the study; one full re-randomization refers to the collection of re-randomizations within *all* pairs.

Table 7. An example of re-randomizing the treatment conditions assigned to two participants within a block.

Participant	**Original Treatment Conditions**	**Re-randomization #1**	**Re-randomization #2**
A	Annotated	*Annotated*	*Non-Annotated*
B	Non-Annotated	*Non-Annotated*	*Annotated*

A randomization distribution lets one view results that could have been obtained if each participant would have received their observed technology score, regardless of what group (annotated or non-annotated) they were assigned to. Thus, it lets one view the variation in random assignment, though under the assumption of the null hypothesis – that the annotated and non-annotated statistics unit have no effect on the integration of technology (i.e., technology scores). Thus, an exact randomization distribution (given this study's design) requires careful consideration of the numerical results under all possible rearrangements of technology scores within a given block. If I sorted participants into four pairs and one triplet (two of which received the annotated statistics unit), then there are two possible ways to randomize the treatment conditions within each pair. Within the triplet, there are three ways ($3\ choose\ 2 = 3$), giving a total of 48 ($2^4 \times 3 = 48$) possible ways of re-randomizing the treatment conditions within all of the blocks.

In the design, I accounted for the three teacher characteristics I expected would have a large impact on the way participants received the treatment conditions by sorting participants into blocks; these were teachers' awareness of the use of technology, awareness of the use of inquiry, and statistical content knowledge. I was primarily interested in investigating the difference in mean technology scores between teachers

receiving the annotated and non-annotated statistics unit, after controlling for these three teacher characteristics.

Although I had no particular interest in comparing the mean technology scores across blocks (the four pairings and one triplet), it would have been important to compare the treatment difference *after* possible differences across the blocks were accounted for. To do this, I would have calculated the difference in means within each block (which is the difference in technology scores for a block that consists of a pair of participants) and made inferences using the average of those differences (Figure 25).

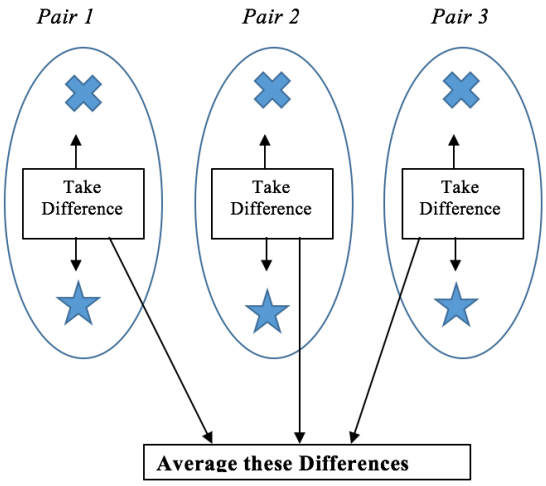


Figure 25: Taking the difference in technology scores within each block (pair) and then using the average of those differences. An "X" represents a participant who received the annotated statistics unit and a "star" represents someone who received the non-annotated statistics unit.

Even though the randomization distribution would be one method for making inferences from the data, I *may* have gained additional insight using a regression approach to make inferences, particularly in light of participants dropping out of the study. Not all of the participants implemented the lessons, meaning some of the blocks

are incomplete. That is, I did not have technology scores for each participant within a block. Multiple linear regression (MLR) models the mean of the technology scores as a function of several explanatory variables and its results are an approximation of the results from the randomization distribution. However, due to the nature of the response variable (technology scores) and the small number of participants in the study, I am aware of the concerns of making inferences from a regression analysis.

MLR analysis makes several assumptions; these include the assumptions of linearity, independence, constant variance, and normality. To check the linearity assumption (examining the relationship between the response and explanatory variables), I would have constructed a matrix of scatterplots of the technology scores versus the technology, inquiry, and statistical content knowledge measurements, coded by treatment condition. In this plot, I would investigate whether the plot of the response means against the explanatory variables was reasonably approximated by a straight line.

With the independence assumption, one examines whether the errors can be modeled as independent. This is often assessed by considering the response variable. In other words, would it be reasonable to suspect technology scores (response variable) from one participant are related to technology scores from another participant? Some examples to consider are whether two or more teachers teach in the same school district, attended the same college for their undergraduate degree, or use the same resources in their classes (e.g., using activities from Illustrative Mathematics), and so on. The design of this study excluded inviting teachers from the same district to participate, and based on

the background of the participants, it was reasonable to suggest there was not a severe violation of the independence assumption.

Next, I would create a residual plot to check the assumption of constant variance, although this assumption is difficult to assess with small sample sizes; this is a scatterplot of the residuals versus fitted values. I would investigate whether the points in this plot were randomly scattered around zero, indicating this assumption has not been severely violated.

The normality assumption assumes the residuals are normally distributed. One can also check this assumption by examining whether the distribution of the response variable (in this study, technology scores) within each group can be approximated by a normal distribution. While the severity of violations of normality are difficult to check with small sample sizes (for instance, with a QQ-Plot), in this study, the technology scores are integer-valued which is inconsistent with the support of a normal distribution. With this understanding, I recognize the normality assumption is severely violated, regardless if all 11 participants had completed the study.

For the regression analysis, I would have used two different MLR models; the first being a model incorporating blocks as a categorical explanatory variable. This would be a direct approximation of the randomization distribution because it accounts for the blocks in the model. The next MLR model I would have used is a further approximation of the randomization distribution. Here, I would have removed blocks from the model and treated the characteristics used to create the blocks (technology ranking, inquiry ranking, and statistical content knowledge) as continuous explanatory variables. This type

of model is useful if one is interested in explaining the relationship between the mean technology scores and the technology ranking, inquiry ranking, and statistical content knowledge.

The inferences from the randomization distribution would have provided a basis of comparison for this study's data to what I would have inferred using the regression models. This comparison may have shed light on how similar (or dissimilar) these methods are for making inferences.

Qualitative Analysis of Classroom Observations. Quantitative technology scores do not describe the entire story about *how* and *why* participants integrated technology and used the annotations during the statistics lessons. Therefore, I also analyzed classroom observations (from Lessons 3 and 4) qualitatively, purposely selecting two participants' classroom observations. I used this qualitative analysis to gather additional information on the efficacy of the annotations. The two participants I selected were chosen primarily for convenience, as they had completed the unit by the end of the 2015-2016 academic year and provided me with complete sets of recorded classroom observations. While these two participants were not paired together in the experimental design, they did share similar qualities, in that both were subject to the annotated statistics unit and were teaching a section of AP Statistics, in addition to Algebra II or Mathematics III, during the 2015-2016 academic school year. Yet, both implemented the annotated lessons (and used technology) in different ways.

I used the field notes and classroom observations to qualitatively analyze the two participants' classroom observations. I coded these documents based on the methods

described by Savin-Baden and Major (2013); the authors suggest the following phases in analyzing qualitative data: characterizing, cutting, coding, categorizing, converting and creating. Although I did not transcribe the recorded classroom observations, I did complete detailed notes (the field notes and observation protocol) for each classroom observation. I parsed these data several times, and indicated general themes of the data. Next, I "cut" the data into manageable and meaningful coded segments. The initial codes I searched for were representative of various instructional actions (e.g., asking questions, forming groups). Some of these initial codes were combined with others to create a new instructional action. I then separated these actions into sub-codes, representing diversion from or adherence to the annotations and lesson plan. Once all of the codes were delineated, I converted the instructional actions into dominant themes; among the themes I expected to emerge are the types and manner of technology use, its users, and the level of guidance given to students during class.

Initial and Exit Surveys. Because I administered these surveys online, all responses were instantly accessible in written form without transcription. I did not code these surveys; rather, they were used to corroborate actions I observed during the classroom observations. Furthermore, I compared participants' responses from the initial and exit surveys; I compiled participants' responses to each of the six questions for ease of comparison (the format thereof is presented in Table 8). This established any changes in participants' views and helped explain why participants implemented the lesson plans and annotations (if applicable) the way they did.

Table 8. Organizing a participant's responses from the initial and exit surveys to make comparisons.

Question	**Initial Survey**	**Exit Survey**
1. Productive use of technology.		
2. Role of exploration and inquiry.		

Interviews. I used the exit interviews to obtain more information about the implementation of lessons and to "fill in the gaps" from any incomplete classroom observations. These interviews also confirmed behaviors I observed during the classroom observations. After each interview, I created a log to organize and summarize topics, as well as additional insights about the implementation, and structure and norms of a participant's classroom. The interviews also shed light on the efficacy of the annotations and necessity for future revisions. Ultimately, these interviews were especially necessary to address the third research question (focusing on the different characteristics of teachers, their classrooms, and implementation of the annotations and lesson plans).

Teacher Background and Experience Survey. To clarify the study's scope of inference and help address the third research question, this survey provided additional data to establish the similarities and differences between participants. I summarized the multiple-choice and short responses and tallied them to obtain a sense of the overall characteristics of the study participants. The open-ended responses were also used to corroborate actions I observed during classroom observations.

Quality and Rigor of the Research Study

In any study, there is a need to address the quality and rigor of the research. There is much literature regarding the quality of a study and how to achieve this goal (Savin-Baden & Major, 2013). Quality, indeed, connotes different things in qualitative (e.g., trustworthiness) and quantitative (e.g., reproducibility) research, and there are many different strategies one may use to ensure quality in research. Regardless of the type of study, the researcher should discuss the relevance of the methods employed and the integrity of the results.

A researcher's own personal stance – that is, his or her perspectives and beliefs – can shape the research study. I have considered my own influence on this study by establishing my background and personal beliefs about using technology and inquiry in a mathematics and statistics classroom. I also conveyed how the frameworks (TPACK and theoretical) used during this study influenced how I perceived and interpreted the data; this was to demonstrate that the results were shaped by the participants and not my own biases.

I have purposely chosen a careful construction of this study's design, paying particular attention to the setting and characteristics of participants in addition to particular choices made in the design of the study (such as ranking participants instead of blocking them into broader categories). I used multiple sources of data, methods of analysis, and investigators to provide corroborating evidence and verify the findings of the study (triangulation). That is, I obtained different but complementary data on the use of technology and inquiry to teach statistics and mathematics; I used classroom

observations, interviews, and surveys to substantiate my findings. I used quantitative methods of analysis together with qualitative methods of analysis to support the strength of my findings. I used my advisors and peers as both listeners and questioners to publicize my thoughts on the data and obtain guidance for its analysis. Multiple outside researchers (those familiar with the study and those not) helped confirm the revisions of the TPACK rubric when evaluating sample classroom observations. Moreover, they verified the technology and inquiry rankings of my participants and the resulting homogeneous pairs/triplets of participants I created.

I employed multiple rounds of evaluation to assess the classroom observations with the TPACK rubric. After all of the lessons were evaluated, I took time to confirm the scores' consistency within and between lessons. With each technology score, I wrote a brief explanation for recording that score. Additionally, I created concrete descriptors and examples within the TPACK rubric to reduce its subjectivity. Overall, these strategies helped me justify the technology scores recorded for participants.

Multiple instruments used throughout this study underwent a process of revision and were not created from scratch. I developed the statistics units in collaboration with other instructors and then implemented and revised them before disseminating the statistics units to participants. I adapted the survey questions from previous surveys (e.g., Mudzimiri, 2012) and tested them beforehand with local high school teachers to anticipate responses from potential participants. I then further refined these survey questions to more accurately reflect the intentions of the research study.

CHAPTER 4

RESULTS

Introduction

In this chapter, I will present results from the qualitative and quantitative analyses of data gathered throughout this study. The goal of this study was to investigate how specially annotated lesson plans influence in-service high school teachers' use of technology when teaching statistical concepts. In Chapter 3, I discussed the methodology, which combines quantitative and qualitative measurements and methods of analysis to investigate this broad goal. Briefly, high school teachers were randomly assigned to receive an annotated or non-annotated statistics unit. I observed each participant teach four statistics lessons and evaluated their integration of technology, recording five technology scores (one for each of the four components of the TPACK rubric and a sum of those four scores) for each participant and each lesson.

First, I identify the final group of participants for this study, as not all 11 participants previously described were able to implement the statistics unit. Next, I focus on presenting the quantitative results using all of the final participants' technology scores. These quantitative technology scores, however, do not convey the entire story of how and why participants used the annotations. Therefore, I follow the quantitative results with more in-depth results from a qualitative analysis of two selected teachers' data. I conclude this chapter with an overall summary of the results.

Final Group of Participants

The attrition rate for participants in this study was high (Figure 26). At the end of the recruiting phase, 70 high school teachers had completed the enrollment survey and stated they were teaching Algebra II or Mathematics III during the 2015-2016 academic school year. Of these 70, there were 30 who completed the initial survey and expressed interest in participating. Of these, 28 had access to a class set of tablets or computers (a necessary criterion) and were invited to participate in the study and receive a statistics unit. By the time I disseminated the statistics units to participants (in December 2015), 11 of these 28 teachers remained in the study. Unfortunately, only five of these 11 teachers were able to implement all four statistics lessons by the end of the 2015-2016 academic school year; the remaining six indicated they would implement the lessons at the start of the 2016-2017 school year, by mid-September 2016. Yet, of these six (as of November 2016), none of them had implemented a statistics unit. As a result, the final group of participants in this study consisted of five high school teachers. The data collected from these five participants were quantitatively and qualitatively analyzed.

I randomly allocated treatment conditions (annotated or non-annotated statistics unit) to the 11 participants by the end of December 2015, first sorting these 11 participants into as similar of groups as possible (four pairs and one triplet). By mid-September 2016, at least one teacher per pair had implemented the four statistics lessons; however, no teachers within the triplet did so by that time (Table 9). Overall, three teachers implemented the annotated statistics unit and two teachers implemented the non-annotated statistics unit.

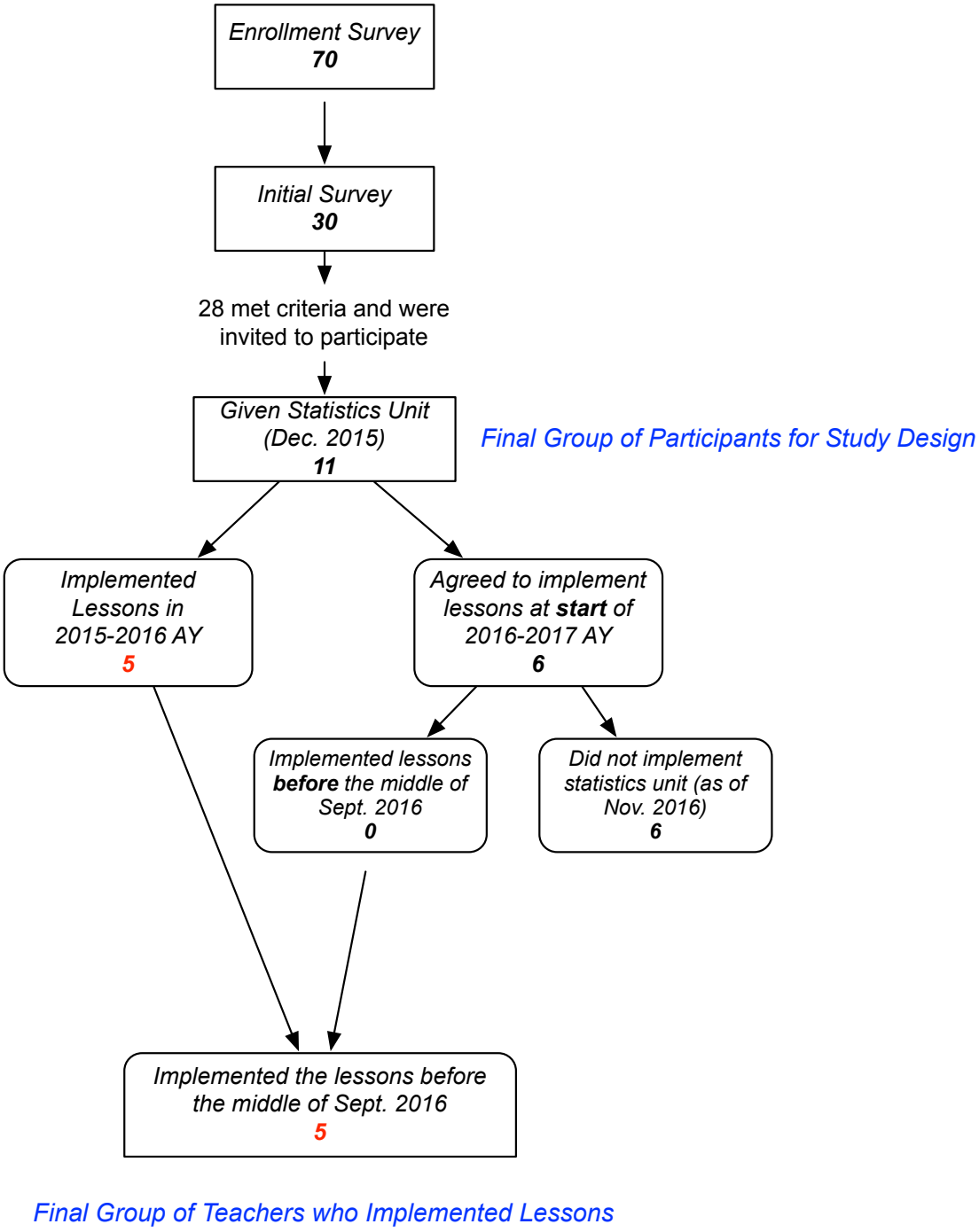


Figure 26. The attrition rate for participants in the research study.

Table 9. A breakdown of the five participants (out of the 11 who received a statistics unit) who implemented the four statistics lessons. The shaded rows correspond to the final five participants.

Participant Number	Treatment Condition	Block	Implemented Lessons?
001	Annotations	1	**Yes**
002	No Annotations	1	No
003	No Annotations	2	No
004	Annotations	2	**Yes**
005	Annotations	3	**Yes**
006	No Annotations	3	**Yes**
007	Annotations	4	No
008	No Annotations	4	**Yes**
009	Annotations	5	No
010	No Annotations	5	No
011	Annotations	5	No

Background Information
for the Final Five Participants

The final five participants varied in their statistical content knowledge and pre-experiment views on the use of technology and inquiry in a mathematics classroom (Figure 27). Their statistical content knowledge scores ranged from approximately 47% to 93%, with a mean percent score of 68%. Furthermore, these teachers populated opposite ends of the spectrum of possible technology and inquiry rankings, as this included those with the lowest and highest rankings among the original group of 11 participants intending to implement the lessons.

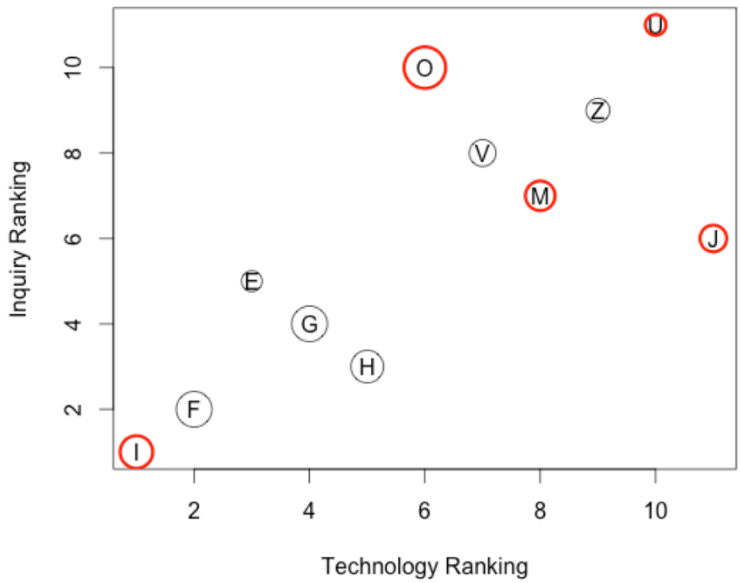


Figure 27. The technology rankings, inquiry rankings, and statistical content knowledge scores of the final five participants (red circles).

All of the final five participants were licensed to teach high school mathematics in their states at the time of this study. During their collegiate study, all five completed an Elementary Statistics course, while two also completed additional, more advanced statistics courses. Their amount of teaching experience varied, but all were considered "experienced" teachers. This group of teachers had taught mathematics in grades 8-12 (as of August 2016) for 27, 9, 18, 27, and 6 years, respectively. Additionally, they had taught an Algebra II or Mathematics III course for 20, 9, 17, 12, and 6 years, respectively. Only two of the final five participants had experience teaching AP Statistics (for 9 and 10 years, respectively) and had attended professional development (PD) workshops focusing on teaching statistics. Additionally, these five participants taught at schools (during the 2015-2016 academic school year) varying drastically in size. The total number of students in these teachers' schools ranged from 65 to 1100 students. The approximate

number of students in their Algebra II or Mathematics III course(s) ranged from 10 to 30 students per class. Furthermore, the size of their mathematics department varied from 2 to 11 mathematics teachers.

Quantitative Results

I originally planned to use a randomization distribution and linear regression analysis to investigate my first two research questions. However, due to the attrition rate of participants, I only had five teachers implement all four statistics lessons. Given this sample size, a graphical analysis of the classroom observations was more appropriate for investigating my research questions. By using this analysis, I was able to obtain more meaningful information about the technology scores than I would have using the more formal statistical analyses I originally planned on using. These, conversely, would necessitate that I summarize all of the data in terms of an aggregate (i.e., averages). This differs from the graphical analysis that allowed me to view all five participants' individual technology scores simultaneously, providing a more complete story of my quantitative data.

In the subsections that follow, I present a graphical analysis and elaborate on the observed similarities and differences in technology scores between participants and lessons. I begin by displaying the individual technology scores, discussing the emergent patterns and investigating the dependence among scores for the four components of the TPACK rubric. Next, I use visual representations of the data to investigate the first and second research questions, first focusing on data from Lessons 3 and 4 and then Lessons

2 and 5. I present additional plots that provide insight into my third research question, seeking to explore the relationship between the technology scores and the three measured teacher characteristics (awareness of the use of inquiry, awareness of the use of technology, and statistical content knowledge). Next, I visually explore the relationship between technology scores and the amount of time participants spent implementing the lessons. I conclude this section with a brief summary of the quantitative results I observed.

Graphical Displays of the Data

I observed five participants teach Lessons 2 through 5; three implemented the annotated statistics unit and two did not. I recorded a technology score for each of the four components of each participant/lesson combination (Figure 28). In this profile plot are two vertical panels; these indicate the type of statistics unit (annotated or not) received. There are four horizontal panels, one for each lesson. A single point in this figure represents the technology score (an integer value between 0 and 4) for a given participant on a given component and lesson; lines connect points belonging to the same participant. The color of the lines and points represent the block (i.e., pair) participants were sorted into prior to random assignment of type of unit. In the end, only one complete pair ("Block 3") implemented all four lessons.

I observed variation in how participants integrated technology during the lessons, both between lessons and between those in the annotated and non-annotated groups (Figure 28). There was no clear difference in technology scores between the two groups; some participants in the annotated group had lower technology scores than others in the

non-annotated group and vice versa. For example, consider the two participants in "Block 3" (teal). Within this pair of participants, I noticed, in Lesson 5, the one who received the non-annotated statistics unit generally obtained higher technology scores than his or her counterpart in the annotated group.

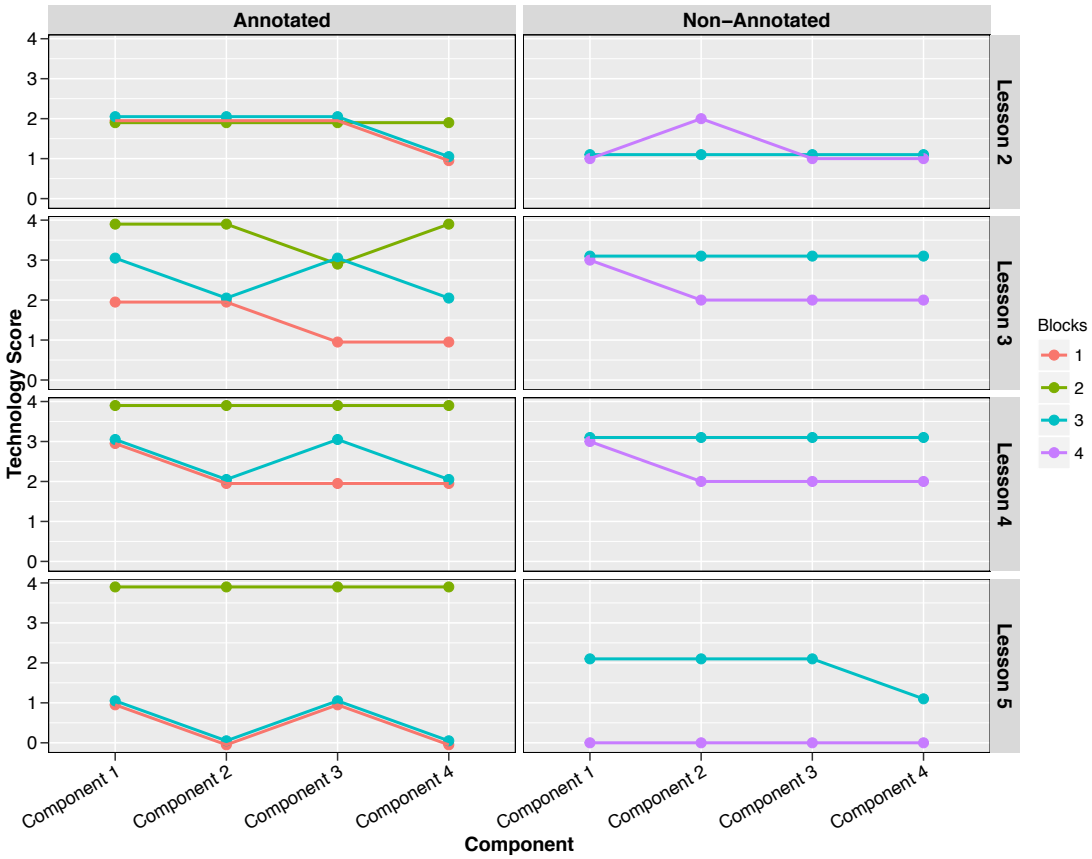


Figure 28. All five participants' *component* technology scores (offset vertically) for each lesson, separated by those who received annotations and those who did not. Lines connect the points belonging to the same participant, and participants sorted into the same block (pair) share the same color.

Upon examining each individual's four component technology scores across each lesson (in Figure 28), I noticed scores for each component within a lesson were very similar within each individual, not varying more than one point in either direction. For

instance, in *Lesson 3*, I recorded the following component technology scores (for Components 1 through 4, respectively) for the participant from "Block 4" (purple) in the non-annotated group: 3, 2, 2, 2. Overall, the separate technology scores from each component did not considerably alter the "overall story" of how a participant integrated technology and inquiry into a lesson. This suggested dependence among the four component technology scores for a given individual. As a result, a single aggregate score, here the sum of the four component technology scores of a lesson, was an appropriate measure to use in the quantitative analysis (Figure 29). This value was an integer that ranged from 0 to 16. A higher total technology score implied a participant integrated technology more effectively. For the remainder of my quantitative analysis, I will be referring to participants' *total* technology scores.

Based on these scores, I noticed the participant from "Block 2" (green) integrated technology more effectively than the four other participants in all lessons. I also observed higher total technology scores in Lessons 3 and 4 than in Lessons 2 and 5. Additionally, there was a steep decrease in scores from Lesson 4 to 5. In the subsections that follow, I address these observations in the context of my research questions.

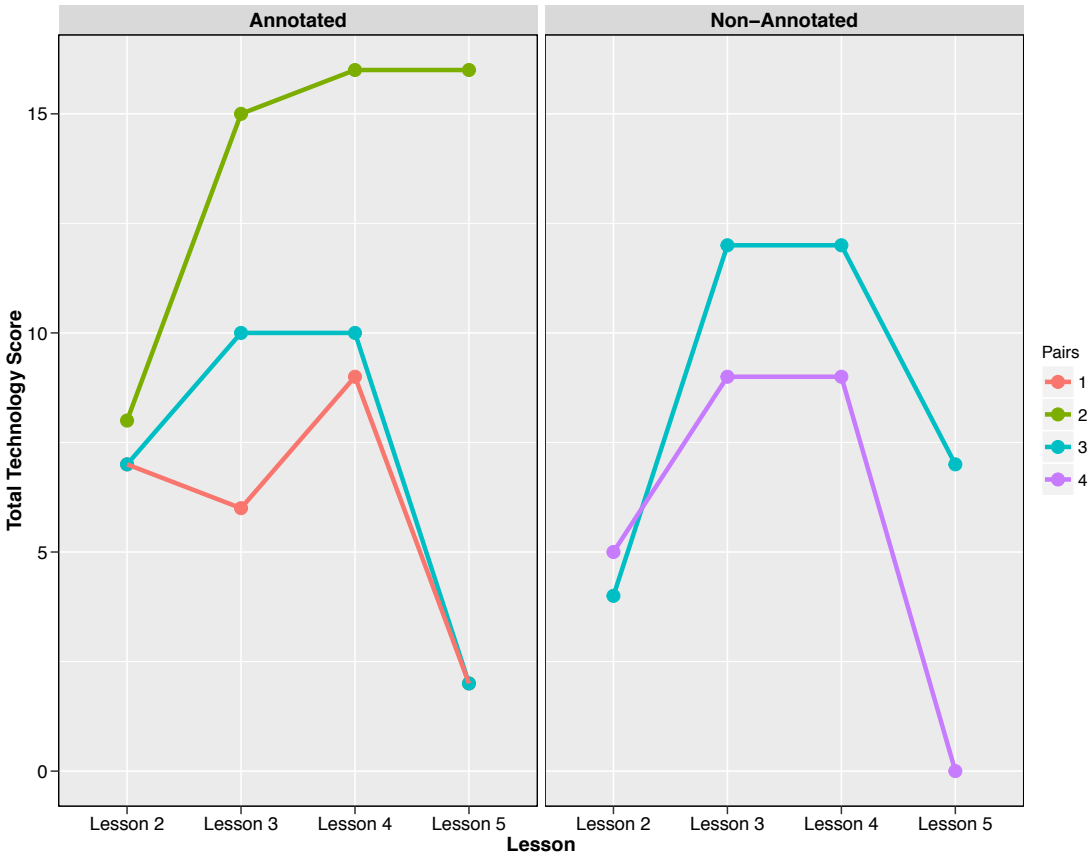


Figure 29. All five participants' *total* technology score for each lesson, separated by those who received annotations and those who did not. Lines connect the points belonging to the same participant, and participants sorted in to the same block (pair) share the same color.

Investigating Research Question #1

The first question of interest pertained to what happened (in a particular lesson) as a teacher implemented an annotated lesson plan. That is, to what extent did the annotations enable a teacher to effectively integrate technology in their instruction of statistical concepts beyond only providing a technology-based activity? All participants, regardless of the statistics unit received, were supplied a technology-based activity in both Lessons 3 and 4. Only those who received the annotated statistics unit were

additionally given annotations for these two lessons. To investigate this research question, I examined the total technology scores from Lessons 3 and 4 (Figure 30).

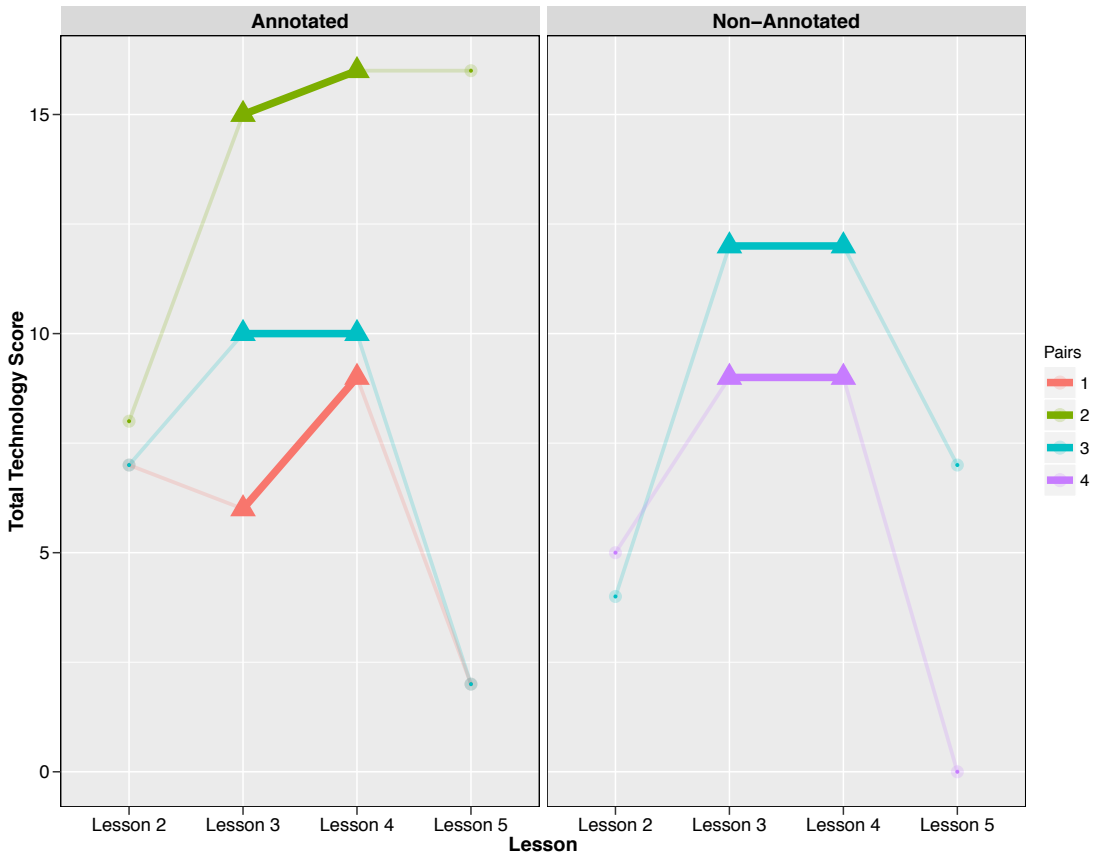


Figure 30. Focusing on participants' total technology scores for Lessons 3 and 4.

Here, I observed most of the participants integrating technology more effectively (corresponding to higher total technology scores) during Lessons 3 and 4 than they did during Lessons 2 and 5. This observation did not surprise me, as *all* participants were provided a technology-based activity during these two lessons. These activities supported the use of inquiry-based exploration (though not all teachers chose to use inquiry-based explorations with these activities) and the use of simulation to teach and learn statistical concepts.

The primary technology used (within the activities) in Lessons 3 and 4 is a web applet that uses simulation to generate re-randomization and re-sampling distributions, respectively. When used appropriately, this applet demonstrates a "dynamic use" of technology that is considered "new and different" from traditional uses of technology such as calculating summary statistics with a calculator. This applet can be used by students to explore new statistical concepts and deepen their conceptual understanding of the material. The "dynamic use" of the web applet also aligns with the two highest levels of the first component in the TPACK rubric, which helps increase a participant's total technology score. I, therefore, expected teachers to use technology more effectively during Lessons 3 and 4 since the technology-based activities supported the effective integration of technology.

I was, however, surprised to observe participants in the non-annotated group integrate technology more effectively than those in the annotated group. For example, the participant from "Block 3" (teal) in the non-annotated group received higher technology scores in Lessons 3 and 4 than two participants (teal and red) in the annotated group. There was no clear difference in total technology scores from Lessons 3 and 4 between those who received the annotations *and* technology-based activity and those who received only the technology-based activity. This was interesting, as the annotations specifically include recommendations on how to use the web applet and activity to their fullest potential. Only one participant (green) integrated technology in ways that aligned with the highest level (in all components) of the TPACK rubric; she had received the annotated statistics unit.

In both Lessons 3 and 4, all but one teacher (red), during Lesson 3 in particular, integrated a dynamic use of the web applet in performing and using a simulation to help students understand the advantages of a randomized experiment; this action resulted in a technology score of 3 or 4 for "Component 1." The participant from "Block 1" (red) encountered difficulties when using the web applet and was not able to use it to generate a sampling distribution. Rather, this participant displayed the end result (a sampling distribution already created using 5000 trials) on a document camera, asking students to use this to complete the remaining questions on the activity. These actions resulted in a lower technology score for "Component 1" (and other components) and a lower overall total technology score.

Before this experiment, three of the participants had never used simulation to teach statistical concepts. At the end of the experiment, all five participants conveyed an interest in using simulation to teach statistical concepts, particularly appreciating the effectiveness of using technology. For instance, one participant wrote

> 1st you said do one of these by hand, 6-10 minutes later they completed it. Then you said do 10,000 of these using the app. Seconds later they had the results. The app allows them to see easily a result that would be difficult w/o the technology.

Another participant commented on how her students were more engaged with the material when they were using the web applet. With support from the technology-based activities, these teachers (who had never used simulation to teach) were now using simulation to teach statistical concepts, a nationally recommended approach.

I also noticed three of the participants (teal in both the annotated and non-annotated group, and purple) received the same total technology score in Lesson 4 as they

did in Lesson 3. This did not surprise me (and confirmed the TPACK rubric was being used consistently between lessons) because I observed each of these teachers implement Lessons 3 and 4 in a nearly identical manner. I found it interesting that the only two participants to improve their total technology score from Lesson 3 to 4 were those who received the annotated statistics unit; yet, the quantitative scores do not relay whether this improvement was a result of the annotations alone. Therefore, I decided to further investigate this qualitatively (presented later in this chapter).

In light of these observations, I cannot conclude that annotations alone enabled participants to integrate technology more effectively. All participants, regardless of receiving annotations or not, demonstrated a more effective use of technology during Lessons 3 and 4 (which contained technology-based activities) compared to Lessons 2 and 5. These observations suggest that creating and providing teachers with technology-based activities (using a data-oriented and simulation-based approach to teach statistics) may not only help teachers integrate technology more effectively in a statistics lesson, but may also guide them to teach statistical concepts in ways that align with the recommendations of the *Pre-K-12 GAISE Report*.

Investigating Research Question #2

The second question of interest centered on how annotated lesson plans may have influence teachers' future decisions about integrating technology in this statistics unit. In other words, did teachers who received annotated lesson plans gain knowledge applicable to a future statistics lesson? To analyze this, I examined the *change* in total technology scores from Lesson 2 to 5 (Figure 31).

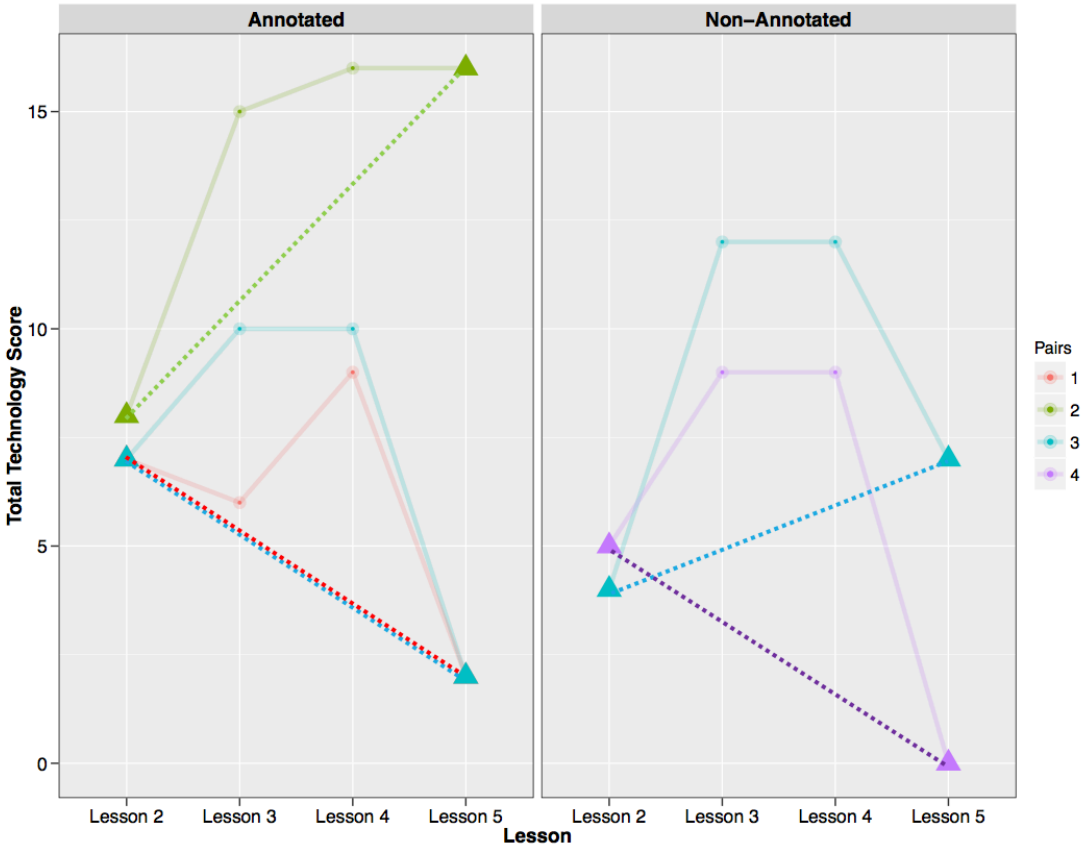


Figure 31. The change in total technology scores from Lesson 2 to 5.

There were no clear differences when comparing the change in technology scores from those in the annotated group to those in the non-annotated group. Only two of the five participants (green from the annotated group and teal from the non-annotated group) displayed improvement in their use of technology from Lesson 2 to 5. The three remaining participants decreased in their total technology scores, all by five points. These observations were the opposite of what I was expecting. I anticipated that participants who received the annotations would be more inclined to use the web applet in Lesson 5 and integrate technology more effectively than they did in Lesson 2 (in which the calculator was the primary type of technology used).

The only two teachers who displayed a more effective use of technology in Lesson 5 compared to Lesson 2 used the web applet in some capacity. The teacher from "Block 2" in the annotated group (green) had an increase of eight points in her total technology score. She was the only teacher who created a technology-based activity and had her students use the web applet to set up, visualize, explore, and discuss the data *and* use a computer simulation to compare mean scores from two treatments. The other teacher from "Block 3" in the non-annotated group (teal) had an increase of three points. She and her students *only* used the web applet to set up, visualize and discuss the data. She and her students did not perform a simulation but instead relied on graphing calculators to compute a test statistic and corresponding p-value. The three other teachers (who decreased in their total technology score from Lesson 2 to 5) rarely used technology as a cognitive tool. They primarily lectured, giving an overall description of the steps used to compare two treatments (such as stating the hypotheses and computing a test statistic).

This decrease in scores from Lesson 4 to 5 and the decision to not use the web applet to perform a simulation may have been a result of participants not receiving a technology-based activity for Lesson 5. Multiple participants asked me to provide instructions and activities to use the web applet for Lesson 5. However, I purposely chose to withhold any activities because of the design of my experiment.

Overall, there was *less* variation in how Lesson 2 was implemented compared to the implementation of Lesson 5. In my design, I purposely choose these two lessons as my "pre" and "post" measures of how participants integrated technology before and after

receiving and implementing annotations and/or technology-based activities in Lessons 3 and 4. I wanted Lessons 2 and 5 to be as similar as possible in their difficulty level and applicability to use technology to teach these concepts. While this is difficult to do given the nature of moving through a unit, these two lessons were more comparable (in difficulty level and applicability to technology) than Lesson 1 (*Randomness*) and Lesson 5 (*Comparing Two Treatments*) were. Overall, the participants generally had a more difficult time implementing Lesson 5 than Lesson 2.

Many of the participants taught Lesson 2 (*The Normal Distribution*) similarly, receiving similar technology scores; they all appeared comfortable teaching the material and, on average, spent the most time implementing this lesson. This content has typically been covered in many Algebra II and Mathematics III courses, as it often appears in a "Probability and Statistics" chapter at the end of textbooks used in these courses (e.g., Larson, Boswell, Kanold, & Stiff, 2012). Moreover, many teachers recognize this topic as assessment items on standardized tests. There is also an abundance of high school resources available online covering this topic, as one participant stated, "I could find a lot of information [about the normal distribution] on the internet" when asked about the resources she used during this unit.

During this lesson, the majority of participants chose to use graphing calculators, applying the *normalcdf()* and *invNorm()* features to compute various probabilities and percentiles. Others made use of calculators to compute test statistics (via arithmetic expressions) and then accessed *z*-tables to compute various probabilities. All participants primarily focused on introducing, explaining, and applying the empirical rule. They

continually asked their students to state the various percentages within a normal distribution (Figure 32).

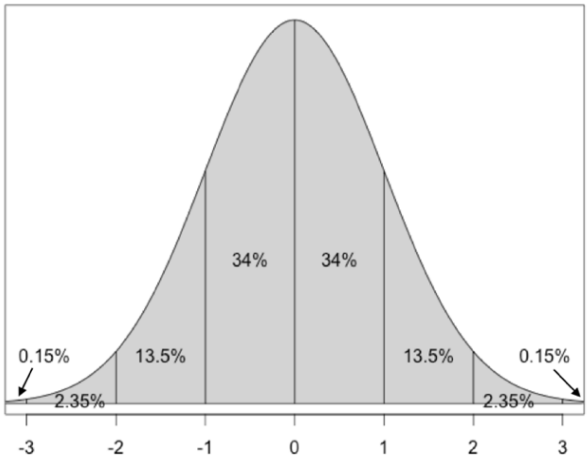


Figure 32. A figure many participants asked their students to fill in with percentages using knowledge about the empirical rule.

The fifth lesson focused on comparing two treatments, a topic that has not typically been incorporated into many Algebra II and Mathematics III courses. This was reflected in the variation of technology scores recorded for Lesson 5 and the drastic decrease in scores for all but one of the teachers from Lesson 4 to 5. On average, participants spent the least amount of time implementing this lesson, and not all of them appeared as comfortable teaching this material as they did teaching the material in Lesson 2. In fact, in an interview, one participant expressed hesitation, stating

> It [Lesson 5] was okay. In hindsight, what I would have to say is that I think that the things I understood well, my kids enjoyed and did well, and the things that I did not understand, then the kids didn't do very well.

When I investigated the change in total technology scores from Lesson 2 to 5, I paid particular attention to the one complete pair of participants I retained for this study

(Figure 33). Within pairs of participants, I expected the one participant who received the annotations to have a higher (positive) change in scores from Lesson 2 to 5. This was not what I observed from my participants in "Block 3" (teal). Still, there was a similar pattern in total technology scores (across the four lessons) and approximately similar technology scores between these two participants. This suggests my blocking strategy (ranking and sorting participants into pairs) may be a useful strategy to employ in future studies provided I am confident about retaining pairs of participants.

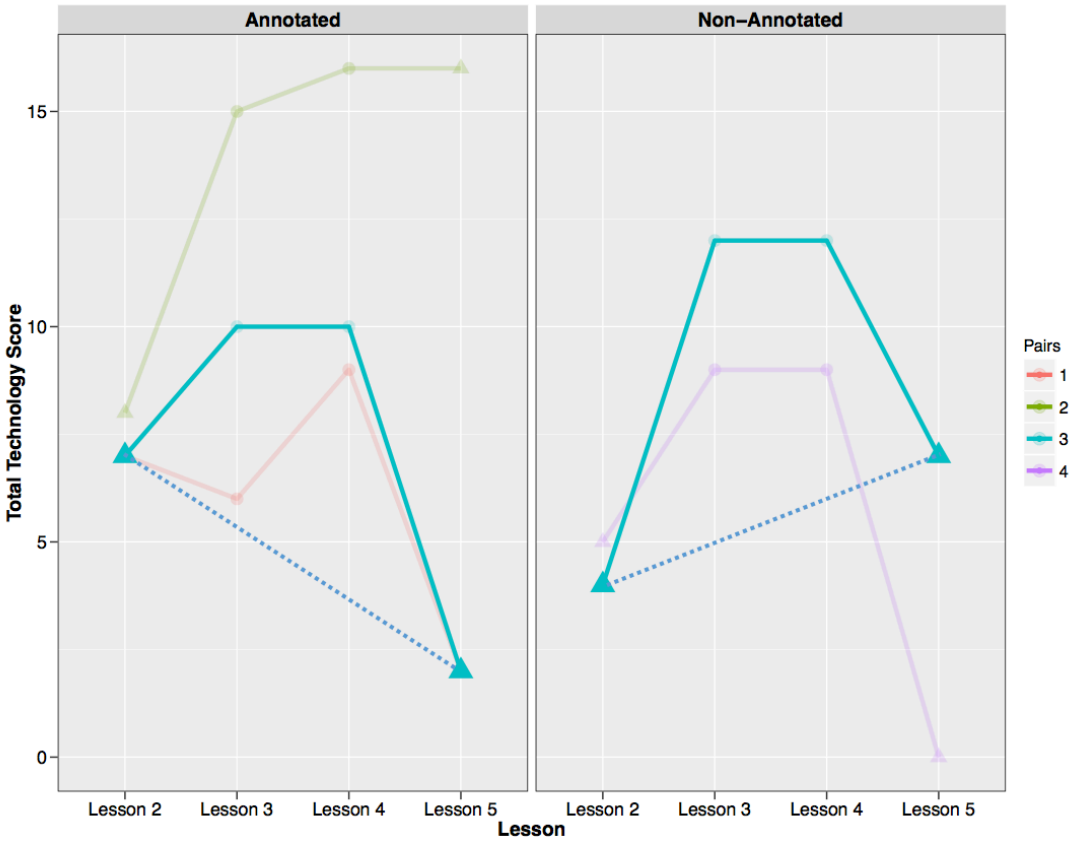


Figure 33. Focusing on the one complete pair of participants (teal) who remained in the study.

Investigating Research Question #3

Morris and Hiebert (2011) discussed how variance in implementation of annotated lesson plans can be influenced by a number of elements (such as differing instructional goals). Therefore, prior to observing teachers implement the annotated lesson plans, I expected they would do so differently, leading to my third research question. This question investigated how particular qualities of teachers may relate to their use of annotated lesson plans as a means to effectively integrate technology in the instruction of statistical concepts.

Prior to participants implementing the four lessons, I collected data about their awareness of the use of technology, awareness of the use of inquiry, and statistical content knowledge. Therefore, I first examined how these three teacher characteristics were related to total technology scores (Figures 34, 35, and 36, respectively). The size of the points in Figures 34, 35, and 36 correspond to participants' awareness of the use of inquiry (inquiry ranking), awareness of the use of technology (technology ranking), and statistical content knowledge score, respectively. Larger points correspond to higher the inquiry rankings, technology rankings, and statistical content knowledge scores, respectively.

I begin by investigating how a teacher's *awareness of the use of inquiry* may be related to his or her total technology score (Figure 34). In this plot, the larger points indicate a higher inquiry ranking. For example, the participant from "Block 2" (green) had the largest points indicating that she was ranked highest in her awareness of the use of inquiry in a mathematics classroom. Conversely, the participant from "Block 4"

(purple) had the smallest points, as she was ranked lowest in her awareness of the use of inquiry.

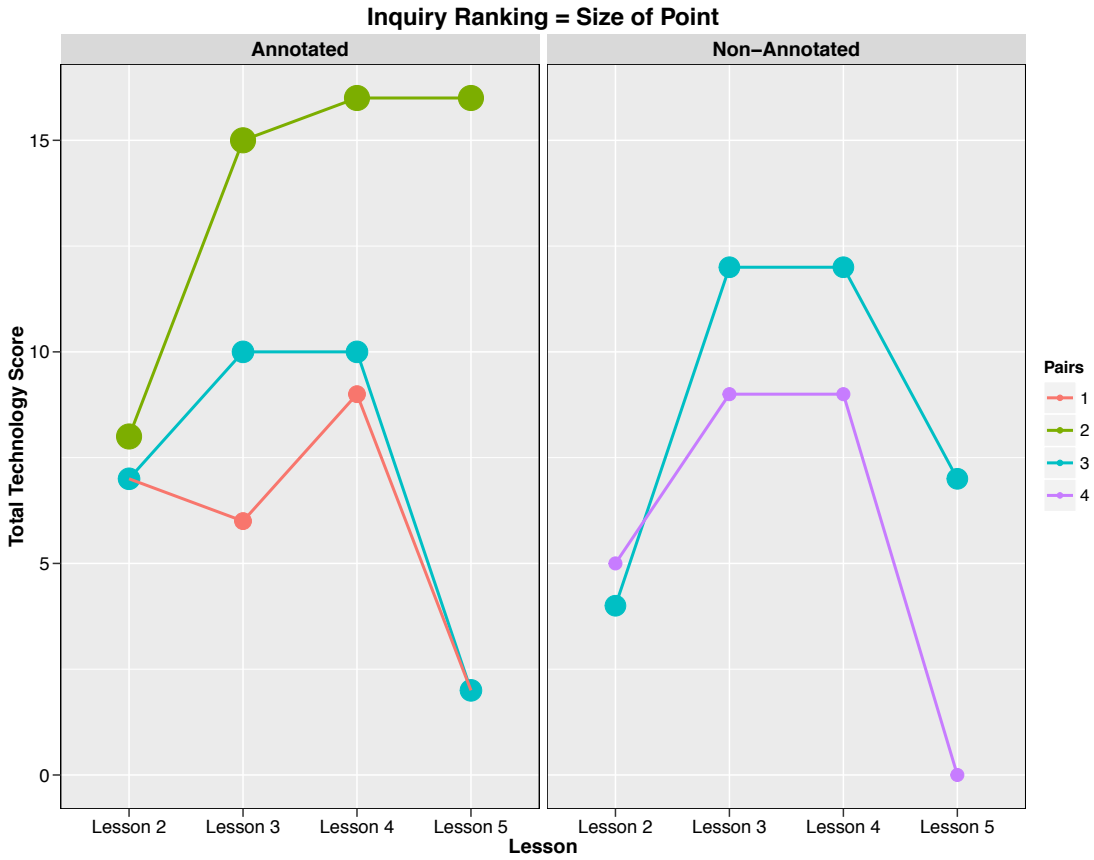


Figure 34. Participants' total technology scores for each lesson. The size of the points corresponds to the participant's inquiry ranking. Larger points indicate higher inquiry rankings.

Looking at all five participants' total technology scores, I observed that participants with generally scores were ranked higher in their inquiry, as well. This suggests that a teacher's awareness of the use of inquiry may be associated with their integration of technology, and perhaps their use of annotations. The later point is something that I further examined in my qualitative analysis (discusses later in this chapter). I also observed that the participant from "Block 2" (green) had the highest total

technology scores (across all lessons); she was also the teacher with the highest inquiry ranking and happened to be in the annotated group. As a result, the treatment condition (annotations versus no annotations) was confounded with a teacher's awareness of the use of inquiry (and also his or her awareness of the use of technology and statistical content knowledge). This is something that I tried to control for in my design, but with the lack of complete pairs, I was not able to. I would have liked to had more complete pairs in my study. In particular, if I retained the green participant's counterpart (in the non-annotated group), I would have been able to make effective comparisons between their scores, after accounting for their awareness of the use of inquiry in a mathematics classroom.

Next, I investigated how a participant's *awareness of the use of technology* may be associated with their technology scores (Figure 35). In this figure, larger points correspond to a higher technology ranking. Once again, the participant from "Block 2" (green) who had the highest total technology score, was also ranked highest in her awareness of the use of technology in a mathematics class. I did not observe the same relationship between participants' technology ranking and their total technology scores (as I did with their inquiry rankings in Figure 34). For example, the participant from "Block 2" (teal) in the annotated group had the lowest technology ranking; yet, this participant integrated technology more effectively than participants from "Block 1" (red) and "Block 4" (purple) (who were ranked higher in their awareness of the use of technology).

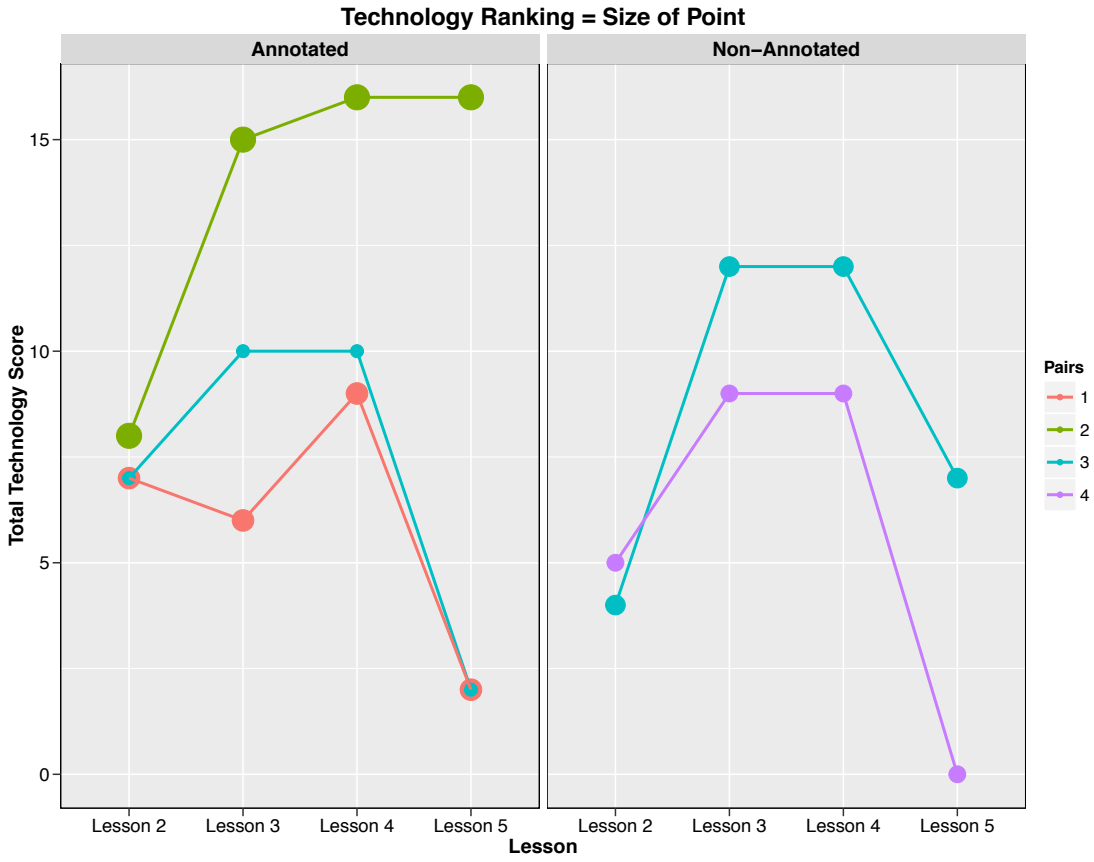


Figure 35. Participants' total technology scores for each lesson. The size of the points corresponds to the participant's technology ranking. Larger points indicate higher technology rankings.

Lastly, I investigated how participant's *statistical content knowledge* may be associated with his or her total technology scores (Figure 36). Here, larger points correspond to a higher statistical content knowledge score. From this, I noticed that participants with higher statistical content knowledge scores did not necessarily integrate technology more effectively than those with lower statistical content knowledge scores. This was seen by comparing the red points to the teal points in the annotated group.

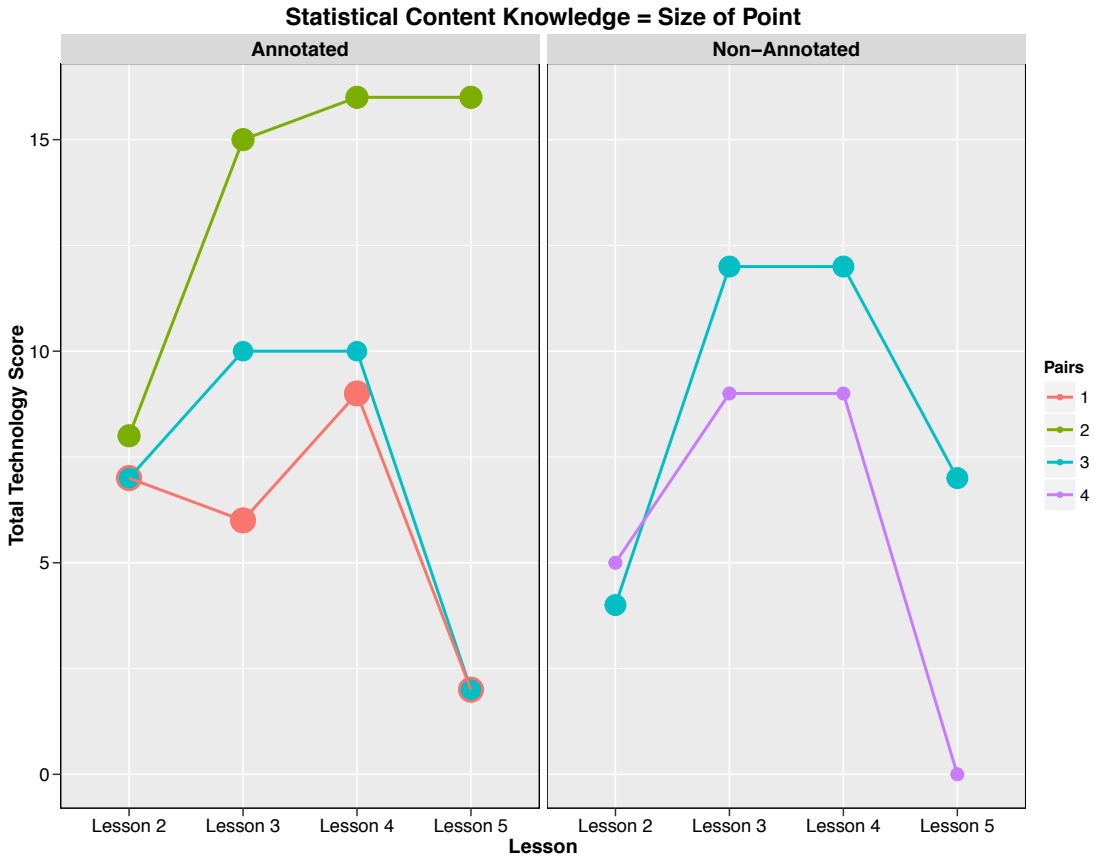


Figure 36. Participants' total technology scores for each lesson. The size of the points corresponds to the participant's statistical content knowledge scores. Larger points indicate higher statistical content knowledge scores.

Figure 34 suggests that a participant's awareness of the use of inquiry was associated with their total technology scores. Inquiry based explorations take time to implement in a mathematics/statistics classroom. During the classroom observations, I noticed the participants were implementing the lessons in differing amounts of time. Therefore, I wanted to explore the relationship between a teacher's awareness of the use of inquiry, time spent implementing the lessons, and his or hers total technology scores.

The recommended maximum amount of time (in minutes) to implement Lessons 2 through 5 were 75, 100, 100, and 75, respectively. Lessons 3 and 4 came with a

recommendation of 100 minutes because of the additional technology-based activity. Often, I use a technology-based activity to teach both Lessons 2 and Lesson 5, using more time in the process. My participants used different amounts of time to fully (or partially) implement each lesson (Table 10). Overall, most of the participants used approximately one 50-minute class period day (or less).

The participants spent the most time, on average, implementing Lesson 2. I was not surprised by this, as they appeared more comfortable teaching the normal distribution than any other topic. However, from the perspective of difficulty and students' understanding, the topics of Lessons 3 through 5 require more amount of time compared to Lesson 2.

Table 10. The approximate amount of time (in minutes) each participant spent implementing Lessons 2 though 5.

Participant	Lesson 2	Lesson 3	Lesson 4	Lesson 5
001	75	50	50	20
002	140	140	140	130
003	50	50	50	50
004	55	55	140	55
005	75	75	50	50

Overall, most participants spent, on average, less than 60 minutes implementing Lessons 2 through 5 (Figure 37). In this figure, I plotted the average time participants spent versus their average total technology scores for Lessons 2 through 5. Participants who received the annotated statistics unit are represented by circles, and participants who received the non-annotated statistics unit are represented by triangles. The size of the plotting characters corresponds to the participant's inquiry ranking; larger plotting characters represent a higher inquiry ranking.

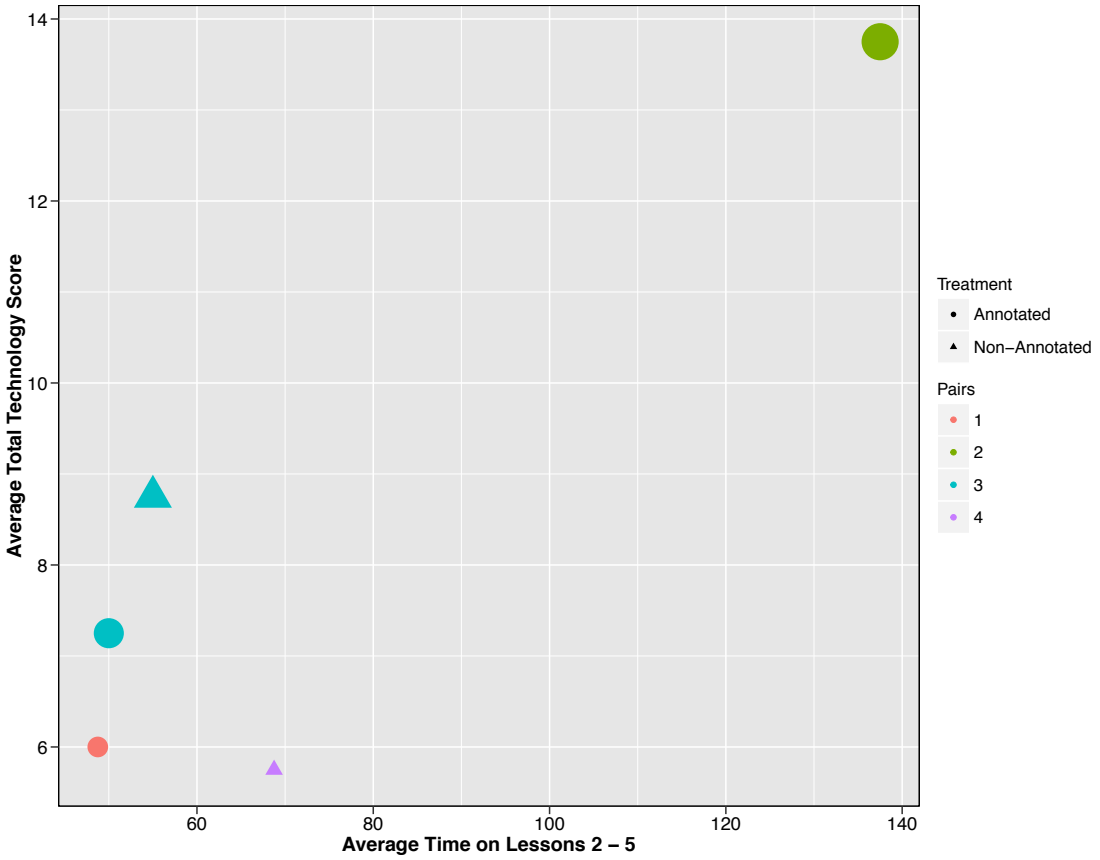


Figure 37. Participants' average time spent implementing Lessons 2 through 5 versus their average total technology score. Those who received the annotated statistics unit are represented by circles and those who received the non-annotated statistics unit are represented by triangles. Larger plotting characters correspond to higher inquiry rankings.

Only one teacher used, on average, more than 100 minutes to implement Lessons 2 through 5; this teacher was the one who integrated technology most effectively and had the highest inquiry ranking. However, other participants who were ranked relatively high in their inquiry (for instance, the two teal participants) did not necessarily use more time to implement Lessons 2 through 5. Furthermore, the green participant was also one of two participants (purple and green) who had block schedules, meaning they had approximately 70-minute class periods. Additionally, these two teachers were using an integrated curriculum and teaching Mathematics III. The three remaining teachers (who

used less than 60 minutes, on average, to implement Lessons 2 through 5) had regular 50-minute periods and were teaching Algebra II under a traditional curriculum. These differences in schedules and curricula may provide a teacher more (or less) opportunity to implement a statistics unit well. This idea is further described in my qualitative analysis.

Summary

This graphical analysis provided insight into how teachers were integrating technology during this statistics unit. There was variation in how teachers implemented the statistics unit within and between the annotated and non-annotated groups. Although I observed no clear difference in technology scores between those who received the annotated statistics unit and those who received the non-annotated statistics unit, there were other interesting findings.

For all participants, I observed higher total technology scores for Lessons 3 and 4 than for Lessons 2 and 5. This suggests that just providing participants with a technology-based activity (using simulation) helped them effectively use technology to teach statistical concepts. Furthermore, these activities introduced a few of the participants to the concept of using computer simulation to teach statistical concepts (a current recommendation from many national organizations). Overall, these technology-based activities helped participants follow the guidelines and recommendations set out in the *Pre-K-12 GAISE Report*.

I observed more variation in total technology scores in Lesson 5 than Lesson 2. Furthermore, only two of the participants (one from the annotated group and one from the non-annotated group) integrated technology *more* effectively in Lesson 5 than in Lesson

2. Within the one complete pair, the participant who received the non-annotated statistics unit had a greater change in technology scores from Lesson 2 to 5 than her counterpart in the annotated group. However, the similarity (in pattern) of their scores across the four lessons indicated the blocking strategy worked well.

Lastly, the figures incorporating the three teacher characteristics suggest that a teacher's awareness of the use of inquiry in a mathematics classroom may be positively associated with their technology scores. It appears that the participants who were more aware of the use of inquiry integrated technology more effectively throughout the statistics unit. Additionally, most teachers, on average, spent less than 60 minutes implementing Lessons 2 through 5. While it is not desirable to teach this statistics unit within a week, the participants' decisions (regarding implementation time) may have been a direct result of the type of curriculum and bell schedule employed by their school.

While the technology scores presented in this section provided insight into how participants integrated technology during Lessons 2 through 5, they do not illustrate a complete picture. Teaching statistics, particularly with technology, is a multifaceted skill that uses a wide range of knowledge; a single total technology score does not describe *why* or *how* technology is being integrated during these statistics lessons. To further investigate why these differences in implementation occurred and what teacher qualities are associated with a participant's integration of technology and use of the annotated lesson plans, I performed a qualitative analysis. I purposely chose two individual teachers who received the annotated lesson plans and analyzed their classroom observations (from Lessons 3 and 4), surveys, and interview responses.

Describing the Two Individual Cases

I selected two individual teachers, Karen and Robin (pseudonyms), to explore the different ways in which they used the annotations; this was in order to better understand and describe the results of my graphical analysis. Both teachers implemented the *annotated* statistics unit yet received different total technology scores (Figure 38). They taught in a Rocky Mountain state, and, in addition to teaching an Algebra II or Mathematics III course in the 2015-2016 academic year, they were both teaching an AP Statistics course at this time. Using information provided in the teacher background and experience survey, I offer a description of each teacher along with corresponding classroom structure.

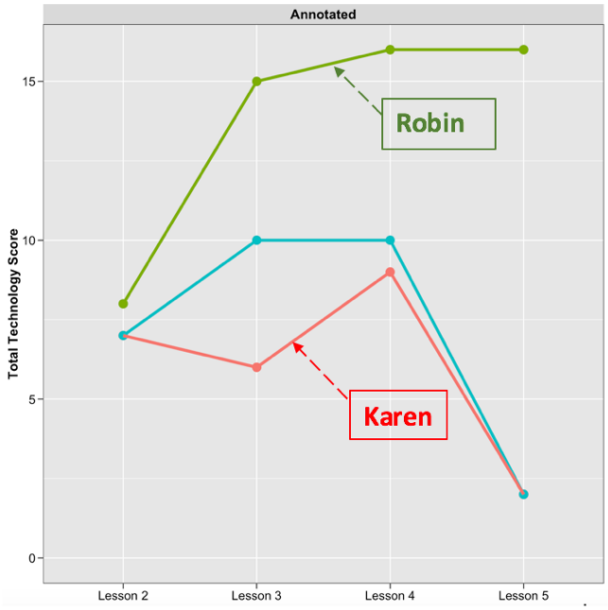


Figure 38. Comparing Karen and Robin's total technology scores for Lessons 2 through 5.

Karen

Karen earned a Bachelor's degree in Secondary Education (Mathematics) and a Master's degree in an unspecified field. In her collegiate study, the highest statistics course she completed was Elementary Statistics. The 2015-2016 academic year was Karen's ninth year teaching mathematics in grades 8-12; during each of these years she taught Algebra II or Mathematics III and AP Statistics. Karen has attended a few regional AP Statistics conferences and a "couple of professional development workshops focusing on technology in the classroom." The only tool she has "been trained to utilize while teaching statistics is the TI-84 graphing calculator." Including Karen, there are approximately eight mathematics teachers in the mathematics department at her school for 640 students.

Karen stated how her school has, within the last two years, begun to introduce and teach statistics in Algebra II courses. The Algebra II class I observed Karen teach consisted of approximately 25 students, and each student had access to a personal Chromebook and graphing calculator. Her class contained individual desks in which all of the students faced the front of the classroom. She had her students work together throughout various parts of the unit, but, due to the placement of the camera, I was not able to see how her students rearranged themselves.

Robin

Robin earned a Bachelor's degree, also in Secondary Education (Mathematics), as well as an Ed.D. in Mathematics Education. During college, she completed an Elementary Statistics course (as an undergraduate) and some more advanced statistics

courses (as a graduate student). The 2015-2016 academic school year was Robin's 28th year teaching mathematics in grades 8-12, and she taught Algebra II or Mathematics III and AP Statistics in approximately ten of these years.

For the past five years, all of the PD she attended focused around "teaching with teams, tasks, and implementing the Teaching Practices from Principles to Actions." Prior to this, Robin participated in summer and year-long PD workshops centered on the use of technology and teaching statistics; for example, she completed an AP statistics workshop from the College Board. During these PD opportunities, Robin learned to use "TI calculators to randomize chances and to do simulations" and throughout the years has learned to "use the calculators and spreadsheets apps like Excel or Numbers to input data, calculate different statistics and to run statistical tests and produce p-values." Robin described the changing role of technology during these PD workshops:

> Particularly with technology, it seemed we first were all amazed at what the technology could do, but lost sight of the fact that students were just following instructions/steps and didn't know what the results meant. However, I was involved with some statistics PD where we actually minimized the role of the technology and learned how to engage students in random sampling so they could see what was happening and what was being controlled. While often tedious and time consuming, it is with these activities that I have learned to help students to understand what statistics and statistical testing is all about.

Robin taught at a smaller school than Karen; Robin's school consisted of approximately 325 students, and there were three teachers (including her) in the school's mathematics department. She implemented the annotated statistics unit in a Mathematics III course. This unit was a natural fit, as the *CCSSM* statistics content standards covered in the unit were already integrated into the curriculum she used. The class I observed contained approximately 20 students, and each student had access to a personal iPad. Her

class was set up for group learning. There were no individual desks in the classroom; rather, tables were positioned together to seat a group of four students (two on each side).

Karen and Robin's Initial Surveys

I used the initial survey to assess all participants' statistical content knowledge and their awareness (pre-experiment) of using technology and inquiry in a mathematics classroom. Both Karen and Robin had comparable statistical content knowledge scores but differed in their awareness of the use of technology and inquiry (Figure 39). Karen correctly answered 14 of the 15 statistical content questions (93.3%), and Robin correctly answered 11 (73.3%). Robin was ranked highest (1 out of 11) in both her awareness of the use of technology and the use of inquiry in a mathematics classroom. Karen was ranked 6 and 10 in her awareness of the use of technology and the use of inquiry, respectively.

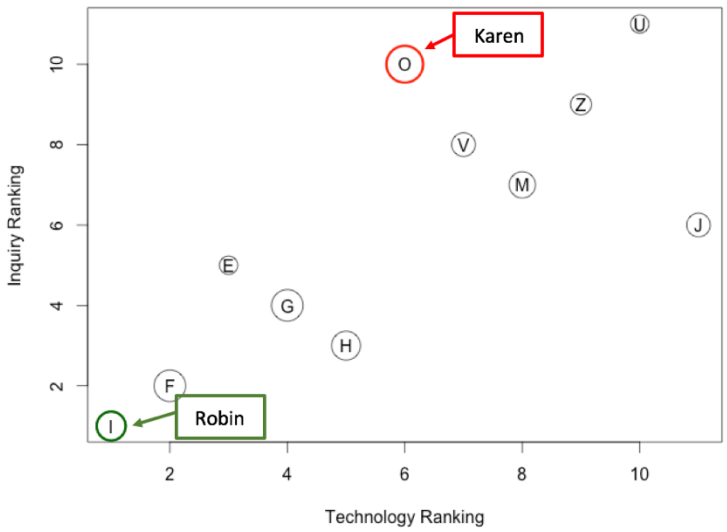


Figure 39. Comparing Karen and Robin's technology ranking, inquiry ranking, and statistical content knowledge scores.

Karen's Awareness of the Use of Inquiry

Of the 11 participants who received a statistics unit, Karen was among the bottom half of the group (with a ranking of 10) in regard to her awareness of the use of inquiry. Before receiving the statistics unit, she expressed how (and why) she relies much more on direct instruction, stating

> Honestly, due to time constraints, the majority of my teaching has primarily consisted of direct instruction, with just a small amount of exploration and inquiry, but I would love to learn new and efficient ways to involve the students more in their education.

These time constraints, in addition to having "so much curricular material in Geometry, and Algebra II that needs to be covered in a finite amount of time" have limited Karen's use of inquiry in her Algebra II classrooms. However, she conveyed using inquiry to teach statistics, stating, "In my AP Statistics course, we do a lot of inquiry and exploration, so I can see it fitting very nicely into the Statistics unit of Algebra II."

Based on Karen's responses, I could not assess *how* and *why*, but rather *when*, she used inquiry and exploration in her class. As she stated, the course she tends to use inquiry and exploration tasks in her AP Statistics course, but not in her Geometry and Algebra II courses. As Karen implemented Lesson 3 (*Observational Studies versus Experiments*), I observed her telling students that she has her students in AP Statistics conduct an experiment, and they [students in her Algebra II course] will get to do that if they take her class.

> When you get to taking AP Statistics that's what your final project is. You get to pick something that's interesting to you and you get to design either an observational study or an experiment and you are going to come up with a hypothesis and then you are going to test that hypothesis.

During my high school internship, in my observation of two AP Statistics courses, I observed that the amount of curricula material stressed in this course is roughly equivalent to other classes, with the bulk being in preparation for the AP Statistics Exam at the end of the school year. These observations suggest that Karen is more receptive to using inquiry and exploration tasks with students who are more advanced in their mathematical and statistical background.

Karen also expressed a desire to "learn new and efficient ways to involve the students more in their education." The key word of *efficient*, in conjunction with her preferred method of teaching in which she uses direct instruction because of limited time, reveal a need to quickly cover all of the material. However, the use of inquiry often requires students to explore and discover concepts on their own and thus can be time-consuming, conflicting with this need. Her responses about using direct instruction aligned with the lower levels of the TPACK rubric, where instruction with technology is teacher-led.

Robin's Awareness of the Use of Inquiry

Robin was the highest ranked with regard to the awareness of using inquiry in a mathematics classroom. When describing the role of exploration and inquiry in the teaching and learning of mathematics in her classroom, she discussed the purpose for using inquiry to teach mathematics along with the students' role.

> My belief is that students learn best when they are engaged in problem solving and thinking rather than just hearing or watching what to do. Therefore, in my classroom students work in teams with team roles (as outlined in CPM [College Preparatory Math]) to keep them on task and working together to make sense of problems and consider different processes. Rather than giving students story

> problems to complete after they have learned an algorithm, a problem is presented first and then used to lay open the mathematics and the algorithmic tools to solve the problem more efficiently.

Robin also considered the role of questioning students, acknowledging a need for her students to drive their learning, asking each other more questions rather than her, the teacher, posing questions to them stating, "While my students are doing a good job of engaging and thinking, one of my goals for this year is to push them to construct questions about their work and to use their mistakes to further their learning."

In her responses, Robin described an inquiry and exploration task she uses in her classroom. This task focused on students working together in teams to collect various data, choose how to graphically represent the data, and formulate questions that could be answered with the data.

Robin's responses revealed an understanding of *why* inquiry is used in a mathematics classroom – to engage students and promote working with others to think through, and make sense of and work through problems with the goal of developing a deeper understanding of the material. Based on her responses and the classroom observations, Robin minimizes direct instruction, focusing instead on enabling her students to take an active role in their own learning. Her explicit goal to have students drive their own learning more through their *own* questioning, rather than questioning from her, corresponds to a shift from the teacher directing the learning to facilitating it. These responses aligned with the higher levels of the TPACK rubric, where students are given an opportunity to explore and experiment with the material to deepen their conceptual understanding of mathematical concepts.

Karen's Awareness of the Use of Technology

Karen was situated at approximately the middle of the group regarding awareness of the use of technology in a mathematics classroom. Karen stated that she uses technology "all the time." The technological tools she uses vary "from [her] desktop computer, to [her] Promethean Board, to [her] Bluetooth connected IPad, to student graphing calculators and Chromebooks." She also conveyed *why* she uses technology in a mathematics classroom, stating

> We live in an age where technology is ingrained into our everyday lives, especially the lives of teenagers. By using technology to teach mathematics, it brings the material to a level that students can relate to. It might mean showing an interesting video that I found, asking the students to do online research on a specific topic, or allowing students to create plots, graphs, and tables that help them visualize the information they are working on.

Students in her mathematics classes primarily use graphing calculators to "enable [them] to create graphs, equations, histograms, scatterplots, etc." Karen and her students use the graphing calculator, in Algebra II, primarily when learning about "linear regression and logarithms." She expressed graphing calculators as the most productive (student) use of technology. As a teacher, Karen identified using the Smart Board, iPad, and Google Classroom as her most productive uses of technology, stating

> The Promethean Board and IPad have allowed me to walk around the room, hand a student my IPad and ask them to work out a problem. This way the student can avoid the embarrassment of walking to the front of the class, standing there alone, and hoping they solving [sic] the problem correctly. Google classroom has allowed me to share lessons and information with students.

Additionally, Karen described how she uses technology to demonstrate procedures and present lessons:

> The IPad and Promethean board allow me to present the lesson and demonstrate how to solve the problems, in front of the class. They also enable me to walk around the room while I am presenting the lesson. This helps me get a better grasp on my students' level of understanding. The same is true when it comes to utilizing the graphing calculators. I will demonstrate the procedures in front of the class on the Promethean Board, and then I am able to walk around and answer questions, help to troubleshoot, and check for accuracy.

Finally, she expressed an interest in using technological tools (other than graphing calculators) to teach statistical concepts, as she said, "I would also like to incorporate more computer-based programs in my Statistics course as well."

While Karen's responses revealed an understanding of how technology can help students learn concepts, the way she described "productive" uses of technology aligned more with using technology as a tool for mere efficiency and demonstration, aligning with the lowest levels of the TPACK rubric. Her statements about the graphing calculator parallel her previous statements of how the graphing calculator was the only technological tool she was trained to use for teaching statistics.

While Karen's responses primarily distinguish the use of technology as a productivity tool, she conveyed an interest to expand her use of technology for teaching statistics. However, her responses and the instructional actions I observed suggest her desire to incorporate computer-based programs in an Algebra II course is grounded in a desire for efficiency. That is, even if she has an additional tool to use in Algebra II, she will likely use it as a productivity tool as opposed to a cognitive tool.

Robin's Awareness of the Use of Technology

Robin was ranked highest in her awareness of the use of technology. She was the only participant to clearly convey two distinct roles of technology.

> There are two roles of using technology in my classroom. Because of the availability of IPads and because our textbooks are online, technology is used in the same way textbooks have been used in the past. Also, assignments are listed online through the Canvas program and grading is completed there. Student [sic] practice skills using the Aleks program and are learning to use technology to complete their homework in a different and hopefully easier way. The other role of technology is to help students learn. While calculators can be used for computations, calculators in my classroom are more often used for checking ideas and theories, creating representations, and looking for patterns and analyzing.

Robin also highlighted the importance of using technology to promote learning, stating "I am always looking for ways that technology can be used to move the learning, rather than just getting answers."

Robin acknowledged that she does "use technology to demonstrate," but when describing the most productive use of technology, she focused on students ultimately using it to explore: "much of it [technology] the students use to explore." Examples of her productive uses of technology included using applications such as Desmos, spreadsheets, applets, and Geometer's Sketchpad to visualize, for instance, how "trig functions change for different values of the unit circle." In her responses, Robin identified a dynamic piece of technology being used, corresponding to the higher levels of the TPACK rubric. Technology, according to her, is a "tool for students to explore and to see the mathematics they are working to understand."

Robin's responses conveyed her understanding of the various roles technology can have in a mathematics classroom; technology serves different purposes (both productivity-based and cognitive) in her classroom. This understanding offers a teacher a considerable advantage for effectively incorporating technology. Recognizing that technology can be used for efficiency in tandem with using it to help students learn can

help teachers decide how and when to use particular types of technology during a lesson. She focuses on allowing students to be the primary users of technology, using it to explore, make connections, and do mathematics. These qualities aligned with the higher levels of the TPACK rubric where students use technology to explore and experiment for new knowledge.

Robin also stated she is looking for additional ways to incorporate technology in her classroom. Her responses suggest her desire to incorporate additional types of technology to better help students learn and understand the material. Technology has drastically changed the way statistics is taught, and Robin addressed this in an interview, expressing how the teaching of statistics has changed since she was a student. This was encapsulated by her recollection of a statistics course she had taken in college: "When I think back to the statistics course I had in college, I am just oh wow, they taught this all wrong." Robin is continually looking for ways she and her students can use technology as a cognitive tool to learn statistical concepts.

Lesson 3: Observational Studies and Experiments

Lesson 3 was the first lesson in the annotated statistics unit to contain annotations and a technology-based activity employing the use of simulation. This lesson centers on observational studies and experiments, explaining the advantages of a randomized experiment. Teachers are recommended to begin the lesson by briefly reviewing categorical and quantitative variables, as both types may be used in the activity. An

activity follows this review, and the lesson concludes with a discussion of the take-home messages.

The activity, *Does Music Help Us Study? – Part 1*, constitutes the bulk of Lesson 3 and addresses all of the learning objectives stated in the lesson plan. The activity begins with three questions aiming to help students recognize and distinguish differences between an observational study and experiment. The remainder of the activity focuses on guiding students to learn about the advantages of randomized experiments. The primary experiment in this activity explores the effect of music while studying. SAT scores are used as the response variable, and students first consider and work with a *quantitative* lurking variable (IQ score). The activity introduces the concept of a lurking variable and asks students to provide additional examples. Next, it presents students with sample data (from 12 hypothetical individuals), instructing them to create cards displaying the names, IQ scores, and SAT prep courses from each of the 12 individuals. Students investigate what happens to mean IQ if they randomly separate the 12 individuals into two groups (treatment and control) by shuffling the cards into two piles, calculating the mean IQ score for each group, and then computing the difference between the two sample means. The activity then asks students to explore what occurs with repetition of this process. However, rather than asking students to do this by hand, the activity directs students to a web applet, guiding them through conducting a simulation that creates many re-randomizations, computes differences in mean IQ scores, and constructs a re-randomization distribution. The remaining questions in this activity focus on students identifying the advantages of a randomized experiment and discussing any differences if

students focused on a categorical lurking (SAT prep) instead of a quantitative one. For this last discussion, I provided additional instructions and questions teachers could use to give students an opportunity to use the web applet again, this time investigating a categorical lurking variable.

The annotations in Lesson 3 provide recommendations for implementing this activity. The first annotation recommends placing students in groups. Given that each student accesses the web applet, the annotations caution teachers that slightly different answers are likely to occur due to way data are generated in the web applet. The remaining annotations are sorted into "chunks" corresponding to groups of similar questions on the activity. These annotations provide various instructional actions like what to do with the students' data, meaningful questions (and possible student responses) to ask students, and possible connections to make between the web applet and cards students shuffled by hand. Additionally, screen shots of the interface of the web applet at each step of the simulation process are provided. Key features (not explicitly clear from the web applet) for teachers to use are also highlighted. For example, one annotation describes how a teacher can "click on a single dot" to display important information about it and how a teacher can use this feature to guide students' understanding about the construction of the re-randomization distribution.

Implementation

Both Karen and Robin implemented Lesson 3, and I recorded five technology scores for their implementation. In the sections that follow, I describe their implementations and provide their technology scores.

Karen's Implementation. Karen used one 47-minute class period to implement Lesson 3; during this class period she did not implement the *entire* lesson, as her class did not complete the entire activity or discuss take-home messages. Moreover, Karen did not complete the remainder of the activity the following day, instead choosing to move forward with Lesson 4.

Karen immediately introduced the concepts of this lesson as something her students in AP Statistics use to complete their final project, telling her students "When you get to taking AP Statistics that's what your final project is." This introduction was followed with a brief review (about 5 minutes) of categorical and quantitative variables. Here, Karen defined and gave multiple examples of each type of variable on the Smart Board. She used the definitions of categorical and quantitative variables provided in the lesson plan, along with the two examples of each variable, and shared six of her own examples (e.g., type of pet). She went through these definitions and examples one-by-one with her students, asking students to stand up if the variable she pointed to was categorical. After this review, Karen went through as much of the activity as time allowed.

Karen and her students spent approximately 20 minutes working through the first four questions of the activity. Karen remained at the front of the classroom, next to the Smart Board that displayed the questions from the activity. She read through each prompt and question on the activity, one at a time, briefly asking students for answers; she then explicitly told students what to write as a solution for each question (e.g., "Write pre-test,

post-test, find the difference.", "You can write down IQ.", and "For 7, write center near zero.").

Next, Karen prompted her students to arrange themselves into groups of two or three. Each group used one Chromebook to access the web applet. Karen continued to read through the prompts and provided each group with pre-made cards. She explained how they were to shuffle the cards and identified computations they needed to make. Karen primarily remained at the front of the class as students shuffled the cards. She did walk around class for a brief moment, but I was unable to view what she was doing.

Students used a graphing calculator to first compute the mean IQ scores of the two groups and then compute the difference. After each group calculated their difference, Karen prompted students to compare with another group. She then discussed repetition and how computers programs can do the shuffling (by hand) process rather quickly, stating

> A good experiment has repetition, meaning it does it several times and then compares those results. So we could reshuffle and re-figure these out, but I think that would take us forever to continue doing that like 20 times or 100 times, so there are computer programs that will do this for you."

At this point, Karen instructed each group to set up the web applet while, at the same time, she displayed the web applet on the Smart Board. She demonstrated and explained what buttons to push to set up the applet. Once it was set up, she explicitly told her students what to write as their answer for a question on the activity, stating, "Write down whatever number you have right here [pointing to where students would find the information on the web applet]." She continued this process throughout the remainder of the activity.

At one point, the activity prompted students to generate and plot many re-randomizations; one problem instructed students to select 5000 more randomizations. When Karen demonstrated and instructed her students to click 5000 more randomizations, the web applet did not respond. One student made the comment, “It says disconnected from server.” At this point, Karen decided to use the document camera (doc cam) to display the solutions (provided with the lesson) on the Smart Board; the solutions contained a re-randomization distribution. She instructed her students to “use this plot” to answer the remaining questions on the activity; however, the answers (based on this distribution) were already being displayed through the doc cam. As students were writing down the answers, the bell rang, indicating the end of class.

This was the conclusion of Lesson 3 in Karen’s class, as the next day was devoted to implementing Lesson 4. The class worked through all but the last six questions on the activity in Lesson 3. They did not reach the main questions of the activity, focusing on the advantages of a randomized experiment (a stated learning objective for this lesson).

To summarize, Karen only implemented a portion of Lesson 3. Both Karen and her students used technology; Karen used the Smart Board, doc cam, and web applet as productivity tools, as Karen's use of the Smart Board and doc cam involved displaying examples, questions from the activity, and the interface of the web applet. The doc cam was used as a backup when the web applet did not respond. The students used Chromebooks to access the applet (though not every student had one) and graphing calculators to compute sample means and differences. Karen mostly remained at the front

of the classroom, providing step-by-step instructions, answers to given prompts, and instructions relating to the web applet.

Robin's Implementation. Robin used two 70-minute class periods to implement the entirety of Lesson 3. She modified the activity slightly by including additional space and tables for her students to organize their thoughts and work, and additional descriptions and parts to problems. She increased the number of times students were asked to shuffle the cards from one to five times and also provided "checking points" throughout the activity, asking students to check their answers and stop for class discussion. Examples of these "checking points" included "STOP AND CHECK YOUR ANSWERS WITH THE TEACHER," and "STOP FOR CLASS DISCUSSION." Furthermore, Robin included entry and exit tickets during the two class periods.

Robin began the first day with an entry ticket focusing on material from Lesson 2 (*The Normal Distribution*). After spending approximately 15 minutes on this, she proceeded to Lesson 3. For about 10 minutes, she went through a review of categorical and quantitative variables, using examples provided in the lesson plan in addition to examples she constructed herself. She also asked her students to come up with their own examples. During this time, students worked in their groups while Robin walked around the class, helping as needed.

The remainder of the first day was spent having students work through the first part of the activity in their groups. Although students were naturally in groups of four (based on their table arrangements), pairs of students were working together on the activity; each student used his or her own iPad. Here, Robin alerted students to the fact

they would each obtain slightly different answers based on how the web applet generated data.

Robin asked students to work through small sets of problems with their partner before either discussing the answers individually with the teacher or as a whole group. Robin consistently walked around the classroom as students worked on the activity, facilitating group discussions and guiding students with appropriate questions; rarely did she provide students with a direct answer. Students worked on the first three questions for about 20-30 minutes before checking in with the teacher and proceeding to use and shuffle the (pre-made) cards, constituting the remainder of "Day 1."

After students shuffled the cards into two groups, they used their graphing calculators to compute the mean of each group and take the difference. In this part of the activity, Robin adjusted the directions, asking students to shuffle the cards, calculate means, and take the difference, and to repeat this process for a total of five times (rather than one). As students were shuffling the cards, Robin set up a GeoGebra spreadsheet and instructed groups to enter the five differences they computed. However, only one group had calculated the five differences by the end of the first day.

Robin began the second day of Lesson 3 with an entry ticket asking students about definitions they were exposed to in the first day of Lesson 3. While students were working on these questions, Robin passed out the pre-made cards. Students completed their tasks following a warm-up; this process took about 20 minutes for all groups to complete. Each group reported their differences and entered them into the GeoGebra spreadsheet Robin had set up. Robin then used GeoGebra to create a dot plot of the

students' data (approximately 40 differences) and projected the plot on the Smart Board. She spent about 10 minutes discussing with the class what they were seeing, prompting students to reflect on what the differences meant (e.g., "What does that [pointing to a difference of 6 on the dot plot] tell us?"). She also asked students to discuss the difference between a negative and positive difference and the meaning of this in the context of the two groups. Robin prompted her students to discuss what they would "hope" the difference would be, leading to a good discussion about why the center of the sampling distribution was approximately zero.

During this discussion, Robin introduced the concept of a simulation: "Now each one of those [the dots] is a simulation trial. It's one shot of trying to randomly assign those scores to … or assign people to those two treatments." She made connections back to Lesson 2 and the normal distribution, asking students to describe the re-randomization distribution (shape, center, and spread). Robin made an explicit point of stating that they [the students] could do this process by hand many times, but questioned why they would want to. This set up the purpose of using technology, as Robin stated, "instead of doing this by hand, technology lets us do this automatically."

For the next 30 minutes of class, students worked through the remainder of the activity. They performed a simulation to answer the questions and investigate the advantages of a randomized experiment. During this time, Robin walked around the classroom, helping groups as needed. At one point, while students were working in their groups, one student had difficulty understanding what they were doing with the web

applet. Robin brought this student up to Smart Board and guided the student through the process, asking her questions as she went through the use of the web applet.

Robin brought the students back together as a class during various parts of the activity to discuss what they were doing. For instance, she asked students what was happening to the re-randomization distribution as they clicked 10, 100, and 1000 more trials. Once students created a re-randomization distribution of at least 5000 re-samples, she again asked students to connect their sampling distribution back to the normal distribution and empirical rule discussed in Lesson 2.

With about 5 minutes remaining in Day 2, Robin brought everyone back together to discuss the last questions emphasizing the advantages of a randomized experiment. Due to the time constraint, her students' responses were brief "Yes/No" responses as opposed to nuanced discussion. The last couple of minutes of class were devoted to a short wrap-up of the lesson and an exit ticket that instructed students to "explain how the bias of a lurking variable can be minimized in an experiment."

To summarize, both Robin and her students used technology during this lesson. Robin used the Smart Board, web applet, and GeoGebra, while the students each used an iPad, web applet, and graphing calculator. These technological tools were primarily used cognitively and by the students. Robin gave her students an opportunity to investigate and explore the advantages of randomized experimentation through the web applet. She also employed the use of GeoGebra to help students understand what one dot on a sampling distribution meant. Robin was rarely at the front of the classroom, instead walking around her class and helping groups of students as needed. She continually questioned her

students' responses, requiring them to provide their own reasoning with deep consideration of context.

Karen and Robin's Technology Scores. I recorded five technology scores for both Karen and Robin (Table 11). The first component, *Conception*, evaluates a teacher's purpose and use of technology and technology-based activities. The second component, *Students*, focuses on a teacher's knowledge of students' understandings, thinking, and learning of statistics with technology. The third component, *Curriculum*, assesses a teacher's knowledge of curriculum and curricular materials that integrate technology for the learning and teaching of statistics. Finally, the fourth component, *Strategies*, evaluates a teacher's knowledge of instructional strategies and representations for teaching and learning statistics with technologies.

Table 11. Karen and Robin's technology scores for Lesson 3.

Participant	Component 1 (*Conception*)	Component 2 (*Students*)	Component 3 (*Curriculum*)	Component 4 (*Strategies*)	Total
Karen	2	2	1	1	6
Robin	4	4	3	4	15

Even though Karen and Robin both received the annotated statistics unit, their technology scores derived from Lesson 3 differed. Karen's instructional actions aligned with the lower levels of the TPACK rubric, as she primarily used direct instruction; she was demonstrating *how* to use technology and telling her students what to write for solutions. Robin's instructional actions aligned with the higher levels of the rubric, as she (and her students) primarily used technology as a cognitive tool, engaging students in *doing* statistics and making connections through inquiry-based learning.

Analysis of Implementation

I used a systematic approach to qualitatively analyze Karen and Robin's implementations of Lesson 3 (Figure 40). Throughout the classroom observations, I saw many instructional actions that arise when one teaches a lesson, such as those related to providing instructions, conversations, technology, student interactions, and so on. From the field notes, I cut out segments corresponding to any broad instructional action I observed. I took all these segments and further sorted them into similar groups of instructional actions. My first thought of these groups of actions revolved around the TPACK rubric. For example, any that were related to technology (or any that were related to students and teachers controlling the tempo of the lesson) were sorted into a single pile. Once I had similar piles of segments, I defined a specific instructional action for each of these piles. However, I noticed I only had one or two segments in a couple of piles, thus I reflected on how these piles might be combined with others. Through this process of "shuffling" segments into piles and piles into similar groups, I arrived at the following five instructional actions, representative of all the segments I cut from the field notes: *technology*, *questioning*, *information-giving*, *pacing of the lesson*, and *content*.

Technology is a broad code encompassing the type of technology used, its purpose, and its users. *Questioning* refers to the degree and content of the teacher's questioning of her students or vice versa. *Information-giving* refers to the type of information the teacher provides her students, including answers, examples, and definitions. The *pacing of the lesson* represents the instructional actions the teacher (or students) employs to control the pace of the lesson; these include actions such as the use

of pre-made cards and grouping students. *Content* refers to the statistical content presented in the lesson and the learning objectives of the lesson.

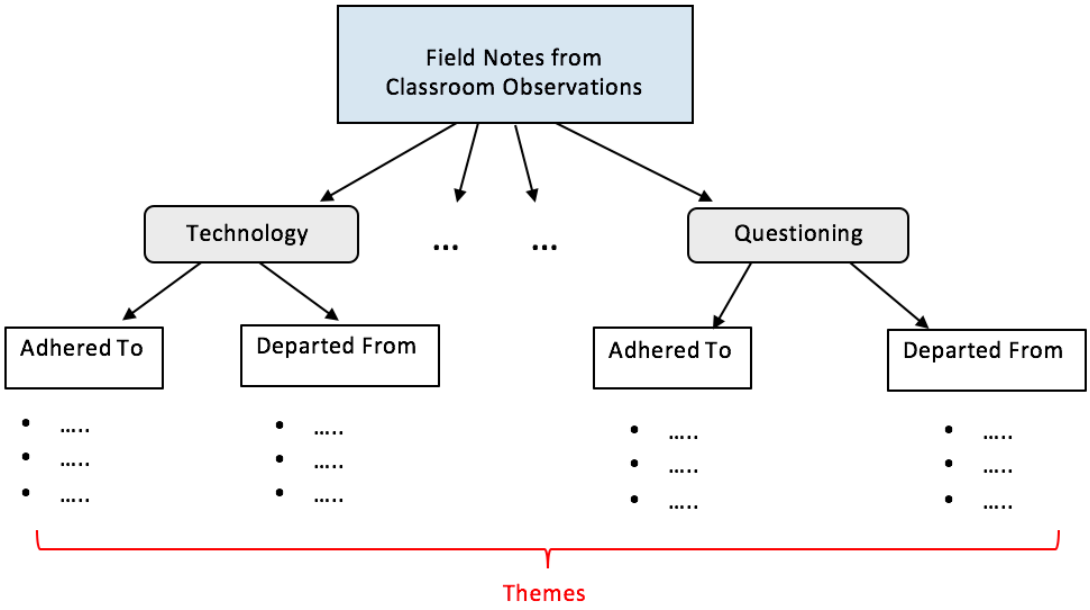


Figure 40. A systematic approach to qualitatively analyze the field notes from classroom observations.

Within each of these instructional actions, I further sorted my data into two sub-codes distinguishing actions that adhered to or departed from the annotations and lesson plans, noting any emergent themes. In the subsections that follow, I discuss the instructional actions I observed, bolding important findings. I first discuss Karen's actions and then Robin's; this is followed by a brief comparison of the two.

Technology. While the *types* of technology **Karen** and her students used (such as the web applet) primarily *adhered* to the annotations and lesson plan, they *departed* in their *use* of technology due to Karen's **management of technology difficulties** and **misconception of how a sampling distribution is created**, thereby **reducing the**

cognitive load of the web applet and activity used during the lesson. Karen and her students, for instance, had difficulties using the web applet, as it was unresponsive after they clicked a button to immediately generate and plot 5000 more randomizations. There may have been issues with the internet connection at her school, as one student told Karen "It [the web applet] says 'disconnected from server.'" In response to this, Karen used the doc cam to project a possible randomization distribution and students used this plot (rather than their own) to answer questions on the activity.

Karen did convey knowledge about the potential pit falls of using technology in her exit survey stating, "…however, it [the web applet] did freeze up from time-to-time, which made it more difficult to impossible to continue. That's the down-side of technology." Her use of the doc cam as a backup demonstrated her ability to quickly adapt when technology fails. However, in doing so, the cognitive demand of the web applet and activity was lowered, as students were no longer given the opportunity to create the re-randomization distribution with the applet (on their own) and explore. Rather, a distribution was presented to them (as is often seen in a textbook) and they used it to answer questions without going through the process of having to create it themselves (with technology). Overall, the web applet was not used to its fullest potential to help students develop conceptual understandings of sampling distributions and the advantages of a randomized experiment.

The difficulties Karen and her students encountered with the web applet, along with the limited amount of time she had available to implement the lesson, appeared to contribute to the fact that this lesson was not implemented in its entirety. In leaving the

activity incomplete, her students did not recognize and discuss the advantages of a randomized experiment. In addition, students were not given the opportunity to discuss and develop their understanding of how the web applet was creating the re-randomization distribution and what one dot on the distribution represented (a strong recommendation included in the annotations). In light of these observations, her students were given an incomplete opportunity to reach the learning objectives of the lesson.

At one point, while Karen was demonstrating how to use the web applet, I observed her relay an incorrect statement about what it was doing "behind the scenes." Karen set up the web applet with the correct data while bringing attention to the sample size of the control group (Figure 41). Pointing to "n = 25", she stated "It did a simulation of 25 times." This suggested Karen did not fully understand how the web applet was creating the re-randomization distribution. In fact, one dot in the re-randomization distribution is created by re-shuffling all 50 participants into two groups of 25 each, calculating the mean IQ score of each group, taking the difference between the two means, and plotting that difference. This description of one dot was included (and strongly emphasized) in the annotations, yet Karen did not correctly convey this particular piece of knowledge stored in the annotations.

Karen's *purpose* for using technology *departed* from the annotations and lesson plan in ways that **reduced the cognitive load of the technology and activity**. She demonstrated how to use the web applet step-by-step. She used the Smart Board for efficiency, displaying the questions from the activity which reduced the time it would have taken to write the questions on the whiteboard. The web applet was not used by

students to *explore* and *investigate* the questions posed on the activity; the problems on the activity primarily served as "drill and skill" type of questions, where Karen's students provided answers with no further justifications. Overall, the technologies used during this lesson were used primarily for **productivity and efficiency**. The ways in which these technological tools were used did not help students understand the statistical concepts.

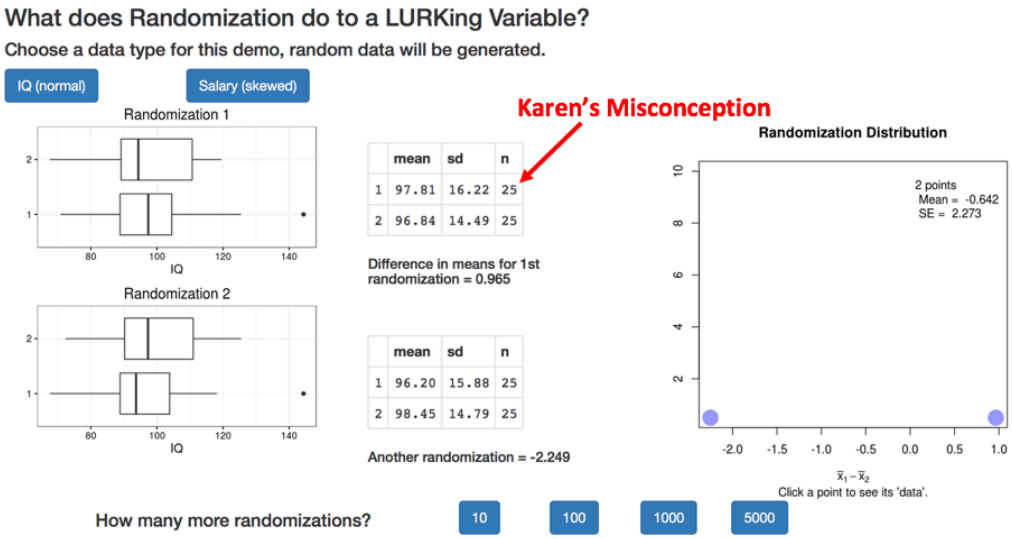


Figure 41. Karen's misconception about what the web applet was doing "behind the scenes" in Lesson 3.

The other participant, **Robin**, primarily *adhered* to the annotations and lesson plan in regard to the *types* of technology used, *who* was using them, and *why*. Robin provided her students an opportunity to use the web applet on their own to explore; she demonstrated its features and use only after students had first attempted to use it. She checked students' understanding with class discussions, additionally using GeoGebra to reinforce what one dot on the re-randomization distribution meant. In this capacity, Robin

often used technology as a **cognitive tool** to **develop students' conceptual understanding**.

Robin also *adhered* to the annotations aimed at helping students make connections between their by-hand work and the web applet; this ultimately helped them **understand what the technology was doing for them**. Throughout the activity, she continually asked her students to describe their actions, first with the cards and then with the web applet. For instance, her additional use of GeoGebra (a recommendation provided in the annotations) allowed her to plot and discuss her students' 40 (by-hand) differences via a dot plot, stating these are the "40 differences from you guys doing simulations." This instructional action explicitly displayed connections between the action of shuffling cards and using the web applet to create the re-randomization distributions. She discussed the meaning of each dot in the dot plot, following a recommendation in the annotations. In an exit interview, she stated

> I love the idea that we actually did it by hand first so I was able to, as I talked about it 'so this is like if we took 50 peoples [sic] names on these cards and we sorted them and we figured out these means. This [the technology] is doing that [by hand cards] for us.

Robin also *adhered* to the annotations when she provided her students the opportunity to use the web applet to explore and investigate statistical concepts. The individual and class discussions during Lesson 3, interwoven with the activity, appear to stem from her experience teaching statistics with simulation and, in particular, its use in the exploration of statistical concepts. In an interview, occurring after she taught this lesson, Robin discussed a statistics book that had "gotten [her] more involved in simulations" and how it incorporates various web applets. She elaborated, in an

interview, that she uses these types of web applets "a lot in [her] AP stats class," and described what she would have liked the web applet in the lesson to include, conveying an understanding of how this web applet could have further helped students understand what was going on:

> I'm thinking the only problem with your app is, you know, you show two simulations which are great and then you are pushing 100 or 1000 or 5000 or whatever you want to create that thing. But it almost would be nice to let it start building and seeing it build them as it goes.

Her students' use of technology *departed* from the annotations and lesson plan when she asked her **students to make connections** to material from the previous lessons. Although the focus of this lesson and activity was on investigating the advantages of randomized experiments, Robin used opportunities during Lesson 3 to ask her students questions about the normal distribution and empirical rule (material from Lesson 2). For example, Day 1 of Lesson 3 started with a warm-up activity where students answered questions involving the use of the normal distribution and empirical rule. By the end of Day 2 of Lesson 3, students once again were asked to make connections to previous material in describing the shape, center, and spread of their own (approximately normal) re-randomization distribution. Additionally, they calculated various probabilities and percentiles for their given distribution. Here, students were doing these types calculations (e.g., 68% of the data lies between what two numbers) by hand or with a graphing calculator.

During these calculations, some students were struggling with deciding when they *needed* to use the calculator (e.g., *normalcdf()* feature) and when they could complete the problems by hand (with the empirical rule). In a few cases, students were using both

methods and exploring on their own with the calculator, asking questions about which answer (0.037 versus 0.039) was correct. Robin used this opportunity to help students understand when to use a calculator and the applicability of the empirical rule, presenting her ability to help students **understand the power of technology** during particular problems.

To compare, both Karen and Robin used similar technologies during this lesson, although the extent to which Robin and her class used technology was more effective for inquiry-based learning. Robin (and her students) used the web applet as a cognitive tool to develop conceptual understanding of statistical concepts. Here instructional actions were also directed at having students understand what technology was doing for them. Karen used technology as a productivity tool and she was the primary user of technology. Her instructional actions were geared toward showing students how to set up and use the web applet, step-by-step. These two teachers showcased very different purposes for using the same type of technology.

Robin's use of GeoGebra to explain the construction of a randomization distribution and her statements (to her students) about what the web applet was doing revealed that she read, comprehended, and used the annotations. Karen demonstrated a misconception about the sample size displayed in the web applet and did not emphasize how one dot was created. These actions suggest she did not fully comprehend (or use) recommendations described in some of the annotations.

Even though both Karen and Robin, for a variety of reasons, departed from the annotations and lesson plan, Robin adhered to all types of technology recommended in

the annotations, only departing in the additional use of graphing calculators to compute various probabilities using the *normalcdf()* feature or empirical rule. Karen's departure (and use of the doc cam instead of the web applet) was a direct result of needing a back-up plan when the web applet failed. These instructional actions removed the dynamic features of constructing a simulation with the web applet, thereby reducing the cognitive load of the web applet and corresponding activity.

Their differing uses of technology and instructional decisions may be traced (in part) back to their underlying beliefs about why and when technology should be used and their comfort levels with simulation and graphing calculators. Karen has acknowledged that she has had training in using the graphing calculator to teach statistical concepts. This aligned with my classroom observations, as Karen confidently dictated to her students how to use the graphing calculator for various statistical procedures (e.g., calculating normal probabilities and percentiles). Her desire to "incorporate more computer-based programs" to teach statistics may suggest that using this type of technology to teach statistics is a relatively unfamiliar approach. In her initial survey, Karen's description of the role of technology in teaching mathematics involved using it so students can "relate" to it; she stated, "We live in an age where technology is ingrained into our everyday lives, especially the lives of teenagers." Her responses align with her desire to use "computer-based programs," as students typically use computers and smart phones daily; still, it is unclear if their use of these technologies is a cognitive use. Based on my classroom observations, Karen used technology to generate and display answers as opposed to exploring and developing conceptual understandings with it. This may be

related to her decision to use direct instruction in Algebra II. Here, she focused on covering as much material as possible in a limited amount of time. Indeed, using technology to explore takes time that Karen may prefer to spend on covering a wide breadth of material.

Robin (through interviews and surveys) has acknowledged that she has experience using multiple types of cognitive technology and has used web applets and simulation to teach statistical concepts before. This was reflected in her implementation as she explained what the web applet was doing and providing students an opportunity to explore with it. During Lesson 3, she did not use technology just for the sake of using it; rather, she sought to help students understand the material and make sense of problems. These actions aligned with her description about the role of using technology to teach and learn mathematics: "I am always looking for ways that technology can be used to move the learning, rather than just getting answers." She reiterated to her students that, when using technology (and particularly graphing calculators), they should always check to see if their answers seem reasonable: "The calculator's going to do all of the calculations for you. You don't really have to do anything other than reason through where things are."

Questioning Techniques. During the lesson, **Karen** did not question her students' reasoning or answers, *departing* from the annotations and lesson plan. She appeared concerned about completing as much of the activity as possible (in one class period) and was **not assessing students' understanding of the statistical concepts or web applet** in the manner described. For example, one particular annotation in this lesson stressed the importance of asking students to explain how one dot in the re-randomization distribution

was created and discussing this concept with them. During one exchange, when students used the web applet to generate one re-randomization and then calculated their difference in mean IQ scores between the two groups, a student obtained a negative difference. Upon seeing a negative difference, the student asked Karen if the difference could be negative. The answer is yes, but Karen replied, "Yup, it can be negative" without any additional elaboration or questioning. She did not ask her students to justify or explain their answers; therefore, it was unclear if her students truly understood, for instance, what they were calculating (and why).

The questions **Robin** asked her students *adhered* to the annotations and lesson plan, **assessing their understanding of statistical concepts and the web applet**. For example, she consistently referred to one dot (either from the class' data or the web applet's re-randomization distribution) and asked her students, "What does that [pointing to one dot] tell us?" In one exchange a student stated, "It's the difference," to which Robin responded, "What does that mean?"

Robin strongly believes students need to show what they understand, stating in an exit interview, "I'm always pushing the, you know, you [the student] got to understand this. I'm not just going to tell you [the student] how to do it." She expects her students to justify their mathematics, and I observed these expectations in her class. She facilitated and guided discussions, asking students to provide and justify their answers.

Statistics, to Robin, is more than just numbers; there is a context behind each and every answer. This belief aligned with her questioning techniques; she often asked students to *explain* an answer (or number) in the context of the problem. For instance, at

one point she prompted students to describe what a difference of −6 meant in terms of the experiment (e.g., the mean IQ of the treatment group was six points lower than the mean IQ of the control group). Her emphasis on context in the classroom was corroborated in an interview, when she stated, in reference to why she emphasizes context,

> I think that's [context] a big one. Because I've been taking this journey towards making sense of math, to me that's the most important math practice standard there is. You need to always make sense of it and I think we have taught for a century or more with just bare numbers that mean nothing. … Those kids that can't memorize, it never makes any sense to them because there is no context to it. … That's why statistics is so great, because it isn't just bare numbers – it really does mean something.

All of her questioning techniques helped her assess whether or not her students truly understood the statistical concepts and what they were doing with technology.

Karen and Robin's questioning techniques differed during this lesson. Karen's limited amount of questions asked (and their types) corresponded to her preference for teaching with direct instruction and limited use of inquiry in Algebra II. She was focused on having students complete as much of the activity as possible. Robin asked students to justify their answers and assessed their understanding of the web applet and. She was concerned with her students' statistical thinking as they were using technology to learn new material.

Information Giving. During Lesson 3, **Karen** *departed* from the annotations and lesson plan by providing students with answers to write down for each problem in the activity. This demonstrated an instructional action where the instructor **provides answers**, thereby **lowering the cognitive demand of the activity**. After going through each individual problem on the activity, Karen would ask students a question and briefly

wait for a response. Upon receiving a response, she would either accept the student's answer or state a different answer they could use. She was explicit in telling students what to write and where, and she provided very clear instructions on where they would find answers in the web applet. For example, while discussing the question – "What differences in students make it hard to get a clear comparison between different music types?" – the following conversation occurred between her and a student.

Student: "How long you study."

Karen: "Write that down."

This process was followed in a similar manner with a few other answers provided by the students. In one occurrence, Karen stated, "What about photographic memories? … You can write down IQ." This appeared to be purposeful, as the activity primarily dealt with student IQ scores in later questions.

Karen's instructional action of **providing answers** appeared to be a norm set in her classroom. During my observations, students did not question her at all, and it appeared natural that she would provide them with answers. This observation was consistent with her responses in the initial survey, where she mentioned her use of "direct instruction" primarily due to "time constraints." Furthermore, her beginning statement in Lesson 3 (regarding how students in her AP Statistics course conduct experiments and observational studies) suggests she views the topic of Lesson 3 as something students will get to and learn if they take her AP Statistics course in a couple of years.

Karen *adhered* to the annotations and lesson plan when she **emphasized definitions of statistical terms**, but she did so in a manner that *departed* from the spirit

of the annotations and lesson plan. This was because the statistical terms (like lurking variable) became the lesson's focus. She did a great job explaining and emphasizing what a lurking variable is (a commonly difficult definition for students to understand), and her students appeared to fully understand this definition. Furthermore, she emphasized this word throughout the activity, stating, for example, "What are some things that might affect your test score that has nothing to do with the music you are listening to?" This provided students with another way to think about what a lurking variable is. She continued to further elaborate on the concept of lurking variables stating, "We said that IQ was a lurking variable so it's not something we can get rid of, but it's something that we could randomize, and that'll help to alleviate the effects of IQ on our results." Although the later part of her statement is incorrect, as one cannot randomize IQ, she appeared to be referring to one of the advantages of a randomized experiment. Karen's emphasis on statistical definitions, however, went beyond what was expected in the activity; definitions were not the primary focus of the activity.

The other participant, **Robin**, *departed* from the annotations and lesson plan when she provided additional information in the form of new instructions and organizational tables in the activity. These departures reflected a sense of consciously **thinking about the needs of her students** and reflecting on what she needs to adjust (in the activity) to accommodate them. She discussed (in an interview) that she took time to read through the lesson plan *and* the annotations, constantly thinking about how it relates to her *own* class.

> It [the lesson plan with annotations] was not something that I felt like I could pick up, read through it and say okay I am ready to go. I really had to work the

> problems, think about, you know, how would the kids interact with this and do all of those things.

Robin did not treat the lesson plan and annotations as a script she needed to follow. Rather, she thought about the material (reading through the lesson plan and annotations multiple times), and interpreted the information provided. This allowed her to transform the stored knowledge for teaching statistics into her lessons. She stated this is a process a "teacher has to go through if you are going to do a good job anyways."

Her decision to have students repeat the process of shuffling cards and calculating difference in means *five* times rather than *one* reflected an understanding that her students needed to go through the (by hand) process multiple times to better understand the meaning of a "trial" and how to build a re-randomization distribution. This instructional decision aligned with her belief that students need to reason through the output from technology; part of this involves understanding what technology is doing "behind the scenes."

Robin also *adhered* to the annotations and lesson plan by **emphasizing definitions of statistical terms**, though this primarily took place outside of the activity, occurring in entry and exit tickets and exam questions. The definitions were not the main focus of Robin's implementation of Lesson 3, but she assessed whether students understood, for example, what a lurking variable is and what happens to it in a randomized experiment.

Karen and Robin demonstrated different ways of providing information to students and emphasizing statistical definitions. Karen focused on definitions and relied on giving her students answers and direct instruction, whereas Robin focused on students

working through the activity and relied on providing students opportunities to explore with the web applet. Robin also included additional information in the activity that would help her students answer the questions and organize their thoughts on their own.

Pacing of the Lesson. In the lesson plan, I suggested teachers allow their students to work through these small sets of problems in the activity in small groups before discussing the problems with the teacher or as a larger group. During Lesson 3, **Karen** taught directly and provided step-by-step instructions throughout the activity and the use of technology. She (instead of her students) read through each of the prompts in the activity. These instructional actions *departed* from the lesson plan and annotations in a way where **the teacher controlled the tempo**, thereby allowing the **teacher to have an active and primary role in students' learning**.

By controlling the pace of the lesson, Karen took control of how her students worked through the activity and to what extent they completed the it. These actions align with her statements in the initial survey regarding her use of direct instruction in Algebra II, primarily due to time constraints. By controlling the pace of the lesson, Karen (and her students) completed more of the activity had the students controlled the pace.

Karen *adhered* to the suggestions provided in the annotations to **manage class time more efficiently** but did so in a manner **to cover as much material as possible**. She pre-made the cards (to be shuffled) ahead of time rather than asking students do this during class. While this action "freed up" some additional class time, the "extra" time was not spent exploring, questioning, or reasoning through the concepts illustrated by the activity. Rather, it allowed the class to complete more of the activity in one class period.

The other teacher, **Robin**, made an instructional decision to walk around the class (helping as needed) as her students completed the activity. Her primary method of teaching revolved around her students being engaged in their own learning. Students were in charge of reading the prompts and questions, and setting up the web applet. The norm of her class was that students work together. She discussed concepts individually with groups and then as a class, but only after all students had the opportunity to work through the problems themselves. These actions *adhered* to the lesson plan and annotations, demonstrating a classroom in which **students controlled the tempo and were engaged with the material**. Here, Robin provided students the opportunity to take an active role in their own learning, even if that meant using more time than originally planned.

Robin also **managed class time more efficiently**, but in a way that **provided additional time to discuss statistical concepts and context.** She made the cards before the lesson to save time; this instructional action afforded her some extra time to use GeoGebra, discuss the construction of a re-randomization distribution, and ask students to provide their reasoning for the majority of their answers in which she stressed the context of the problems.

She also *adhered* to the annotations and lesson plan by asking students to work in groups despite the relative difficulty in management of group work. She acknowledged these difficulties, stating, “That’s the problem with them all working in groups – they all work at different paces.” However, she always had something ready for groups to continue working on after they had reached a certain point: “So, I am waiting for this

group to get to this point, so I got to have something for the other groups to continue doing."

Robin's *departure* from the annotations and lesson plan occurred during the final ten minutes of the lesson, in which, considering the amount of time left in class, she **minimized the importance of the last few questions** on the activity (emphasizing the advantages of a randomized experiment). An annotation discussed the importance of discussing the last four questions on the activity, as they are the main point of the activity. Rather than having a discussion, however, the class went through these questions together with basic "yes" or "no" responses. While Robin did not discuss these questions in detail during the lesson, she did assess students' understanding of these questions in subsequent entry and exit tickets and exam questions.

The disparities in how Karen and Robin controlled the pace of the lesson may be related to their preferred teaching methods and the type of curriculum used. Karen used traditional teaching methods in Algebra II; she used direct instruction and often controlled the pace of the lesson in order to cover as much of the material as possible. Karen expressed a concern about the amount of time she has to cover all of the material in Algebra II. The addition of this statistics unit, in which its content standards have not typically been taught in her curriculum, added additional material to cover in a finite amount of time. This time constraint may have hindered her complete implementation of Lesson 3 and may also explain why Karen sought to control the lesson's pace, providing students limited opportunities to use technology to explore and investigate statistical concepts.

Robin had a non-traditional preferred method of teaching in that she guided students to learn statistical concepts and facilitated discussions, prompting students to take an active role in their own learning; students controlled the tempo of Lesson 3. Robin taught Mathematics III and the statistics content standards from these lessons were already built into her curriculum. Therefore, Robin did not have to fit in extra material. She already had time built in to implement these lessons, favorably impacting how implemented the lesson.

Content. **Karen** primarily *adhered* to the content presented in the lesson plan and annotations. She followed the content outlined in the lesson plan by using the activity. Her *departures* occurred when she did not implement the lesson in its entirety, thereby not covering all of the learning objectives. For instance, because of time constraints, she **did not address the advantages of a randomized experiment**.

During Lesson 3, **Robin** covered all of the learning objectives, *adhering* to the annotations and lesson plan. She made small adjustments to the activity and occasionally *departed* from the content in the lesson. Her primary departures occurred during the warm-up activity which involved questions regarding the normal distribution and empirical rule. These departures revolved around her stated personal goal to **have students make connections to previous material**. Robin made an instructional decision to begin each day of the lesson with a warm-up question asking students to use technology and prior knowledge. For example, on the second day of this lesson, she created a new example of an experiment and asked students to identify the treatment, explanatory variable, and response variable. Furthermore, throughout the activity, she

consistently asked students to "connect this [the current material] back to what we did last week." Much of this was in reference to the normal distribution. She even related the concepts in Lesson 3 to those she used in Lesson 1 (Randomness) and used language she planned to use in future lessons such as, "Is it very likely?" (when referring to one dot on the randomization distribution).

Robin's instructional decision to ask students to make connections becomes clear upon consideration of her exit interview. This is indeed something that she continues to work on in her classroom.

> My biggest problem, especially with these juniors, is that I feel like what I do in class should make really good sense and they're okay. But when I ask them about things, it's like they are not making connections. So I don't know if I am not spending enough time doing that or if it's not that important that they remember that stuff. I am really struggling with everything I've been teaching them all year as to whether we are making any progress.

Both Karen and Robin had similar content knowledge scores from the 15-multiple choice assessment on the initial survey and have experience teaching AP Statistics, implying their *overall* statistical content knowledge is likely to be relatively similar. Karen's instructional actions regarding content primarily aligned with the lesson plan, and Robin diverted from the content presented in the lesson plan to further discuss and ask students to make connections back to the normal distribution. The normal distribution is a topic often taught in Algebra II an dMathematics III and usually appears on standardized tests, which may explain why Robin asked students to discuss and work with additional content (such as the normal distribution) outside the learning objectives of Lesson 3.

Lesson 4: Interval Estimation and Margin of Error

The fourth lesson of the statistics unit (the second to contain annotations and a technology-based activity employing the same web applet used in Lesson 3) focuses on developing students' understanding of confidence intervals and the margin of error. The primary goal of this lesson is for students to understand what an interval estimate and margin of error are, using simulation to develop these in the process. Students ought to recognize the advantages of using an interval estimate over a single point estimate, as well. The first annotation recommends teachers motivate the content through discussion of a real-life study. The bulk of the lesson centers on students completing an activity; this is then followed by a wrap-up of the entire lesson.

The activity in this lesson is preceded by the activity from Lesson 3. Titled *Does Music Help Us Study? – Part 2*. It continues to focus on music and studying but addresses a slightly different research question; students explore a research study investigating whether rural high school students in the U.S. prefer to study with music playing or in silence. Thirty randomly selected (hypothetical) students were asked their preference; 23 expressed a preference for music playing. The researcher of this study is interested in obtaining a "good" estimate of the true proportion of rural high school students in the U.S. who prefer to study with music playing.

The activity guides students through the four-step investigative process prescribed in the *Pre-K-12 GAISE Report*. Students formulate a research question, design a study and collect data, explore and summarize the data, and draw inferences from the data. For convenience, the activity provides the research question to be investigated, proposes a

design, and provides data. The beginning questions encourage students to think about aspects of the study design and data collection. For example, students are instructed to identify the population of interest and the response variable collected. The activity then focuses on briefly exploring and summarizing the data, asking students to identify the sample size, calculate a sample proportion, and think about what would happen if the study was repeated with another group of 30 students. This last question is particularly important and motivates the remainder of the activity.

From here, the activity guides students to understand why a single point estimate (for instance, a sample proportion) does not tell the "whole story," at the same time identifying the need for researchers to know how good their point estimate is. Accordingly, students are introduced to confidence intervals and margin of error. Students conduct a simulation by hand to see how variable the sample proportion they calculated may be. To do this, they shuffle 30 cards (marked with the observed outcomes of the study – 23 "music" and 7 "silence"), pick a card at random, record the outcome, replace the card, and repeat this process until they have 30 outcomes recorded. Then, they compute the proportion of outcomes that were "music." Finally, they replicate this process aided by technology.

The same web applet used in Lesson 3 is again used in this activity for students to conduct their simulations. The activity prompts students to generate many resamples; however, this time, after a re-sampling distribution is created, students focus on exploring confidence intervals with various confidence levels with the web applet. Additional questions are posed to help students understand these topics and to introduce the concept

of a margin of error. Lastly, students are asked to use information from their re-sampling distribution to calculate a 95% confidence interval by hand and to compare their results to an interval generated in the web applet.

The annotations I provided in this lesson plan are similar to those found in the previous lesson. The annotations offer tips, such as what to do with the students' data, meaningful and important questions to assess students' understanding of what the web applet is doing and possible student responses, and connections to make between the web applet and the cards students shuffled by hand. Problems of prime importance are distinguished in the annotations, with emphasis again placed on how one dot is created in the re-sampling distribution. Additional recommendations include emphasis on new statistical definitions and more elaboration on how to interpret confidence intervals in the context of the problem.

Implementation

Both Karen and Robin implemented Lesson 4 similarly to Lesson 3. In the subsections that follow, I briefly describe their implementations of Lesson 4 and any instructional changes from Lesson 3. Finally, I provide their recorded technology scores.

Karen's Implementation. Karen used one 47-minute class period to implement Lesson 4 and (almost) complete the activity. She immediately began with the activity, choosing to not motivate the lesson with any real-world examples. She and her students were finishing up the last problem on the activity as class ended. Therefore, they did not discuss any of the take home messages of the lesson.

While Karen's instructional actions were primarily similar to those in Lesson 3, she did, however, implement slightly different (and more effective) actions in relation to inquiry and technology. Karen primarily stayed at the front of the class, reading the prompts and questions one at a time. She provided answers and directed students on what to write and where throughout the activity. When she used technology, she demonstrated how to use the web applet, step-by-step as students followed along.

One student per group of two or three had a Chromebook for the web applet; Karen's students also used graphing calculators to compute arithmetic expressions (such as a proportion) during the activity. Unlike Lesson 3, Karen and her students did not have any difficulties using the web applet. They conducted a simulation, generated many re-samples, and created a re-sampling distribution. This was a dynamic use of technology employed by Karen and her students that I did not observe in Lesson 3.

Overall, Karen involved the students slightly more during this lesson, occasionally asking them to explain their reasoning. On one occasion, she provided students a brief opportunity to explore with the web applet. She also demonstrated a feature on the web applet that allows a user to click on one dot, displaying relevant information in a table. However, she did not use this feature to emphasize the meaning of one dot.

Robin's Implementation. Robin used two 70-minute class periods to fully implement Lesson 4, modifying the activity slightly. She made slight modifications to the activity; these included emphasizing learning objectives from the lesson, adding additional questions asking students to make connections back to previous material, and

altering questions about investigating the confidence intervals created with the web applet.

Robin's instructional actions were similar to those in her implementation of Lesson 4. This time, she spent even less time at the front of the class presenting material. More time was devoted to student exploration and experimentation with the web applet. Each student used an individual iPad to access the web applet and complete the activity with a partner. Robin walked around class, helping and guiding groups of students as needed. As she helped students, she would question students' reasoning and request further elaboration. Occasionally, she would bring the whole class together to discuss key concepts students were struggling with (to ensure the class was on the same page). She used entry and exit tickets during both days of the lesson, assessing students' understanding of statistical concepts. While she did not initially motivate the lesson with a real-world example reporting a margin of error in a study, she did bring this idea up throughout the lesson, for instance, stating, "You hear this [margin of error] all of the time in polls," followed by an example. Her students completed the activity and had time to work on sample homework problems in class.

Karen and Robin's Technology Scores. The technology scores I recorded for Karen and Robin's implementations of Lesson 4 differed (Table 12). Karen improved in her use of technology and inquiry; during this lesson she demonstrated an improvement in all components but Component 2. She and her students were able to enact a dynamic piece of technology, increasing her score in the *Conception* component. Her additional, though limited, use of technology-assisted exploration improved her scores in the third

and fourth components. Robin demonstrated an effective integration of technology throughout the entirety of the lesson, while aligning closely with *all* the learning objectives – thereby improving from Lesson 3 in the *Curriculum* component.

Table 12. Karen and Robin's technology scores for Lesson 4.

Participant	Component 1 (*Conception*)	Component 2 (*Students*)	Component 3 (*Curriculum*)	Component 4 (*Strategies*)	Total
Karen	3	2	2	2	9
Robin	4	4	4	4	16

Analysis of Implementation

I used the same systematic approach (originally used for analyzing Lesson 3) to qualitatively analyze the field notes from Karen and Robin's implementations of Lesson 4. I was able to sort excerpts from the field notes into the same five instructional actions used for analyzing Lesson 3; namely, *technology*, *information-giving*, *questioning*, *pacing of the lesson*, and *content*. In the subsections that follow, I briefly discuss Karen and Robin's instructional actions. In many cases, the actions were similar to those observed in Lesson 3.

Technology. **Karen** and her students primarily used the web applet during this lesson, *adhering* to the type of technology recommended in the annotations and lesson plan. As opposed to Lesson 3, during this lesson Karen and her students *adhered* to the annotations and lesson plan in their *use* of the web applet. That is, they did not experience any technological difficulties with the web applet and performed a simulation to create their re-sampling distribution. This resulted in the use of a dynamic piece of technology to learn statistical concepts. One Chromebook per group of two or three students was

used to access the web applet; therefore, there were no differences in the number of people trying to access the web applet as in Lesson 3. This suggests that internet issues alone (as opposed to user difficulties) may have caused the technological difficulties encountered in Lesson 3.

Karen's *purpose* for using technology mirrored that of Lesson 3, *departing* from the annotations and lesson plan in ways that **reduced the cognitive load of the web applet and activity**. Karen primarily demonstrated how to use the web applet, step-by-step, with her students following along after each brief demonstration. This time, however, there was one limited exploration opportunity involving the students' use of the web applet, thereby minimally *adhering* to the annotations and lesson plan. This corresponded to a slight shift from the sole use of this web applet as a productivity tool to a cognitive tool. The particular question, in which Karen gave students a brief opportunity to explore with the web applet on their own, was near the end of the activity. It asked students to "Play with the Confidence Level % buttons more to explain: How are the endpoints of the interval estimate related to the colors of the points in the plot?" The keyword, "Play," in addition to having time remaining in class may have encouraged Karen to allow her students to use technology to explore, diverging from her implementation of Lesson 3, when she appeared to rush through the activity. While this instructional action resulted in a very brief moment of students exploring with technology, the primary instruction of technology during this lesson was "teacher-led, student-follow," reducing the cognitive load of the web applet and activity.

During this lesson, Karen conveyed that she and her students were using the web applet because "you [the students] saw this shuffling is slow and takes quite a while. We want to speed up this process, and to do that we can use technology." I also observed a technology demonstration (not observed in Lesson 3) in which a key feature of the web applet was used by Karen and her students. Throughout both lessons, the annotations consistently emphasize the importance of discussing how one dot in the sampling distributions is created and using the "click on one dot" feature. During Lesson 3, neither the teacher nor the students clicked on a dot or discussed its meaning or how it was created, leading me to ponder whether Karen read the annotations or purposely chose to ignore those recommendations. However, during Lesson 4, Karen *did* tell students to click on one dot. She demonstrated this feature showing students the table of information that appears in the web applet when one dot is selected.

I noticed these two changes in instructional actions occurred when these recommendations were **stated directly in the *activity*** as opposed to just the annotations. Including important directions such as "Play with" and "Click on that dot" on the activity itself helped Karen (and her students) use technology more effectively than in Lesson 3. These observations, in tandem with Karen's preferred method of direct instruction in Algebra II, suggest that annotations included directly in the activity itself may have been followed by Karen more than when they were stated solely in the annotations. In addition, the lack of time pressure to complete the activity in Lesson 4 may have also contributed to Karen implementing these two instructional actions.

However, while Karen and her students used the "Click on a dot" feature in the web applet, they did not use it to further discuss how one dot was created; this recommended discussion helps students to understand what the web applet is doing "behind the scenes." I observed Karen relay another misconception about the re-sampling distribution, suggesting she did not understand the web applet at this level.

During the lesson, she displayed a re-sampling distribution (created with 1001 re-samples). She pointed to the center of it and stated "that means that 76% of the students preferred to listen to music while studying" (Figure 42). This statement was her response to a question on the activity that prompted students to state the center of their distribution and explain why it seemed reasonable (commonly difficult for students to grasp). An appropriate response to this question is, in fact, "The simulation is run assuming our best guess at the proportion of high school students who prefer to study with music playing is 0.767." Karen would have been less likely to relay this misconception to her students had she followed the recommendation of discussing one dot's meaning.

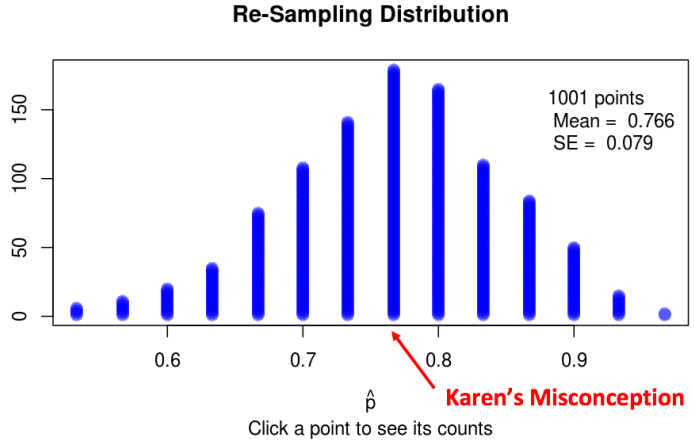
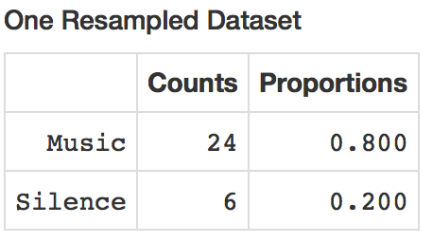


Figure 42. Karen's misconception about what the web applet was doing "behind the scenes" in Lesson 4.

Overall, Karen and her students **did not use the web applet to its fullest potential**. Students went through the motions of using the web applet, not fully understanding what it was doing. They (and the teacher) used their calculators as much as possible and even when unnecessary. For instance, a problem asked students to use technology to generate one re-sample and state the proportion of students listening to music. Prior to this, they were instructed to click on one dot, where a table of information appeared (e.g., Figure 43). The proportion of interest was included in this table, yet Karen instructed her students to calculate the proportion by taking the number (e.g., 24) and dividing it by 30, ignoring the efficiency of the web applet automatically displaying this number. Hence, this was not an instructional action meant to help students recognize information provided in the web applet. I viewed it as a missed opportunity to use the web applet to its fullest advantage and reduce the amount of necessary computation.



One Resampled Dataset

	Counts	Proportions
Music	24	0.800
Silence	6	0.200

Figure 43. One resampled dataset displaying the proportion of interest.

Another use of the graphing calculators involved the calculation of confidence intervals. At one point in the activity, students were introduced to the idea of a confidence interval and why they might be interested in it over a single proportion. Karen motivated this by stating, “So, we came up with 0.77 as our proportion. Well, really a better result would be to say between 0.74 and 0.78, or something like that, and there’s a

mathematical way to find that interval." I did not interpret her use of "mathematical" as referring to a simulation process. Rather, her emphasis on mathematical relayed her inclination towards using a formula to create a confidence interval. For instance, in AP Statistics, students are commonly provided the following formula:

$$\hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$$

This is slightly different than the one provided in the activity to calculate a 95% confidence interval which rounds the value used for z^* and uses the standard error displayed in the web applet ($Estimate \pm 2{\times}SE$). When Karen calculated a 95% confidence interval using the formula provided in the web applet, she incorrectly used the formula $Estimate \pm SE$ instead and wrote the interval as (0.844, 0.69), switching the endpoints. The emphasis on mathematical formulas and graphing calculators, when not needed, suggest Karen was not as experienced in using a computer simulation-based approach to teach statistics.

As in Lesson 3, **Robin** *adhered* to the annotations and lesson plan in the *types* of technology used, the *primary users* of technology, and the *purpose* of using it during Lesson 4. Her instructional actions were nearly identical to those in Lesson 3. In this lesson, she gave students even *more* time to explore and investigate with the web applet. She demonstrated use of the web applet only *after* students had a chance to use it themselves; this action reinforced concepts they had learned. Importantly, she was using technology to **develop students' conceptual understanding** of statistics. She reiterated why the web applet was being used and made connections to the card shuffling students performed. She stated, "We have this nice applet instead of doing the cards which we did

last week. The applet does the same thing, but does it much faster." Her instructional decision to have *every* student use the web applet supported the idea that students would obtain slightly different answers when using the web applet. She purposely wanted her students to recognize the different answers obtained and explain why this was happening.

Throughout the lesson, Robin demonstrated that she had read and understood the annotations. Based on my classroom observations, Robin illustrated that she understood what the web applet was doing "behind the scenes" and how to troubleshoot any difficulties students were having with it. One annotation conveyed a common difficulty students have when using the web applet and a suggestion to state this. Often, after entering data in the web applet, students forget to click "Use These Data" allowing them to continue with the correct data. Before Robin's students could encounter this problem, she reminded students at each table to click on "Use These Data."

Robin stressed the importance of "one dot" to her class and consistently assessed their knowledge of its creation throughout the activity, in entry and exit tickets, and in exam questions. She conveyed a desire to have her students **understand what the technology was doing for them**. Prior to using the web applet, when students were shuffling cards to generate one re-sampled proportion, she asked students for their re-sample proportions, displaying their answers on the Smart Board on a pre-made axis (Figure 44). This differed from her use of GeoGebra for this purpose in Lesson 3 but conveyed that she thought about the best use of technology to demonstrate this concept. When she used GeoGebra in Lesson 3, it appeared to take her longer than she wanted to set up the dot plot of students' data. This time, students were able to see the plot

constructed one dot at a time (instead of all at once as in Lesson 3) as students' proportions were plotted (one at a time) on the Smart Board. Less time was devoted to creating this plot, allowing for more room to discuss "one dot."

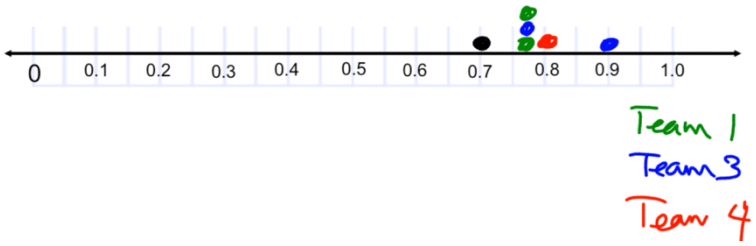


Figure 44. Robin using the Smart Board to display her students' re-sampled proportions after shuffling cards by hand.

Her desire to assess students' understanding of one dot was clear throughout the lesson. She often said, "I need you [the students] to be very clear, so that you know in your mind what each dot is going to mean when you hit 100 [more resamples on the web applet]." In Lesson 3, she primarily stated what one dot meant and how it was constructed. This time, in Lesson 4, she constantly asked her students to state what it meant, guiding their answers to include statistical language. The following exchange occurred when asking the class what one dot on the re-sampling distribution meant.

> **Robin:** "We need to be very clear about this [one dot]. So who can tell me, what does this one dot mean?"
>
> **Student:** "If you did one test … That's …"
>
> **Robin:** "If I did one resampling then …"
>
> **Student:** "The percentage of people that listen to music would be that [motioning to the value of one dot displayed]."

These types of conversations occurred throughout the lesson, becoming more sophisticated as students gained understanding of the concept. I also observed, at times,

her students telling each other to click one dot and discussing its meaning. Her enthusiasm for this idea was effectively relayed to her students.

Robin also demonstrated knowledge about using technology effectively when she adjusted the confidence interval questions in the activity. The original questions prompted students to explore with various confidence levels and asked students to state the number of red and blue points in the re-sampling distribution as these levels were adjusted. Robin altered these questions to focus students' explanation on how the endpoints of the interval estimate relate to the colors of the points. This change conveyed that Robin carefully thought about the questions beforehand, worked through the activity first herself, and considered how technology can best be used to explore statistical concepts.

Robin and her students used **the web applet to its fullest potential** to develop understanding of the concepts presented in Lesson 4. Her students appeared more confident using the web applet during this lesson than in Lesson 3. As they were investigating with it, they asked and reminded each other to use the key features of the web applet; this was not something I observed them doing in Lesson 3. The only time students used their graphing calculator was to compute the sample proportion from the cards. In the web applet, when they were asked to write down the proportion of students in their one re-sample listening to music, they took full advantage of the web applet's proportion display.

Both Karen and Robin used the same types of technology during this lesson, but the extent and purpose to which Robin used technology was more thoughtful and

effective. Robin used technology to help students develop their conceptual understanding of the material, aligning with the annotations. Students in Robin's class were actively engaged with learning concepts with technology on their own but guided by the teacher. Karen did not experience technological difficulties with the web applet and briefly provided her students an opportunity to explore with it. The clear difference between these two participants was Robin's demonstration of implementing recommendations provided in the annotations; she used technology as a cognitive tool to help students *do* statistics. Karen's use of technology was primarily used as a productivity tool, but there was some improvement in its use in some instances. While I was unable to conclude whether or not Karen read the annotations, I saw that she used technology more as a cognitive tool when suggestions were placed directly in the activity.

Questioning Techniques. **Karen**'s questioning techniques were similar to those implemented in Lesson 3, *departing* from the annotations and lesson plan. In Lesson 4, however, she was able to **minimally assess students' understanding of the statistical concepts**. On a couple of occasions, she asked students to briefly elaborate on their answers, allowing for more discussion about their thoughts. Here, the following exchange occurred with the following question: "If the study was repeated with another 30 high school students, do you think you would get exactly 23 again who prefer to study with music playing? Explain."

Student: "No."

Karen: "No, why not?"

Student: "Because they are a different group of people."

> **Karen:** "Ok. So because it's a different group of people this time, so they are going to have different opinions. Every random sample will have a different sample of people, so you are going to get slightly varied results."

Even though the student's justification was brief, Karen did ask the student to *partially* elaborate. Still, she did not ask the student to *fully* elaborate on his response, instead explaining it for him.

In regards to using technology to explore concepts, I witnessed Karen providing her students an opportunity to do this when the question on the activity contained the keys words, "Play with." However, directly asking students to "Explain" on the activity did not prompt Karen in the same way. This was because the questions on the activity from Lesson 3 also asked students to "explain" their reasoning, yet she did not ask students for explanation. This activity, though, was shorter than the one from Lesson 3. This seems to imply that time was a factor in the types of instructional decisions Karen made. She had more time in Lesson 4 (than in Lesson 3) to ask students for some elaboration. However, this elaboration occurred only a couple of times, where the primary use of questioning during this lesson did not ask students to justify their answers.

Robin's questioning techniques revolved around having her **students understand what they were doing with technology** and **assessing their understanding of statistical concepts**, *adhering* to the annotations and lesson plan. Robin demonstrated her desire for students to explain their reasoning while considering the context of the problem (e.g., "How do you [the student] know that?"). What I additionally observed in Lesson 4 but not 3 was *her students* questioning results they were seeing from the web applet. For instance, after Robin displayed a 95% confidence interval on the web applet, I observed a

student ask, "What does that number tell us?" The students were beginning to **question what answers meant in the bigger picture** of statistics without being prompted to do so. Her students were treating answers as more than just numbers; they were attempting to put them into a particular context. Additionally, they began to question the accuracy of results. Robin, at one point, calculated a confidence interval incorrectly using an incorrect estimate; students pointed out the error and explained what values to use instead.

Robin's questioning techniques encouraged students to **take an active role in their learning and understanding**. The types of questions she asked **emphasized the learning objectives of the lesson**. One particular objective is for students to recognize the advantage of using an interval estimate over a point estimate. With this in mind, Robin asked a question *not* included on the activity to drive this point home:

> Now, what do you think that [the confidence interval generated by the web applet] has do with the fact that we are asking a question [the original research question] in the first place? What do you think this [the confidence interval] has to do with answering that question [the original research question]?

She also included an additional question at the end of the activity that reinforced this learning objective: "What is the best estimate of the proportion of high school students who listen to music while studying?" These additional questions *departed* from the annotations and lesson plan, but they did so in a way that **reinforced the learning objectives**.

Karen and Robin's questioning techniques differed for Lesson 4, paralleling most of the techniques they used for Lesson 3. However, both of their questioning techniques improved slightly from Lesson 3. Karen asked more "why" questions, prompting students to partially justify their answers. Robin's push towards asking students to explain their

reasoning and relate their answers to the context of the problem influenced her students, as they began to question what they were doing.

Information Giving. **Karen** *departed* from the annotations and lesson plan by providing students with answers to write down for each problem in the activity. She told students what, where, and when to write, stating, for example, "just write that down for your answer on part d" and "so, for #9, just somewhere on there, your answer is whatever your number is out of 30, and then do that division on your calculator and write that down that decimal." She demonstrated an instructional action where the instructor **told the students what to do and gave them the answers**, thereby **lowering the cognitive demand of the activity**.

When students provided an answer, Karen would either accept a student's answer or state a different answer they could use, even when their version of the answer was correct. This had the effect of **minimizing the students' voice in the classroom**. For example, the following conversation occurred between Karen and her students in response to what the researcher (in the activity) should record when they collect data.

> **Student:** "You write down whether or not they like music playing or not."
>
> **Karen:** "Ok. So what exactly are you going to write down. If they say I like to study in silence what are you going to write down?"
>
> **Student:** "They like to study in silence."
>
> **Karen:** "Or you are probably going to just write silence, right? And if they say I like to study with music, you are going to write …"
>
> **Students:** "Music."

> **Karen:** "So, for number 1, the response that you are going to record is music or silence."

In this exchange, it appeared that Karen wanted to have students write down "Music" or "Silence" as their answers because the activity (later on) did create these nicknames for when students were asked to make cards to shuffle. For example, the cards displayed "M" for music and "S" for silence. However, having students say and write exactly what the teacher has in mind minimizes their ownership of the material and their own learning.

Robin, the other teacher, *departed* from the annotations and lesson plan with her inclusion of additional questions in the activity and rearrangement of problems, **thinking about the needs of her students**. She reorganized the activity in a way that would make sense to her students on the given day they worked on the lesson. After Day 1 of Lesson 4, she made further alterations to the handout to give to them on Day 2.

Robin's *adherence* to the annotations revolved around using more statistical language (e.g., parameter) and interpreting confidence intervals and the meaning of confidence in the context of the problem. This was a recommendation of the annotations. She also allowed students to state their own answers, giving students a **voice in the classroom**. She accepted multiple versions of an answer provided students were either correct or on the right track. In the following exchange, multiple students provided justification for why they may not obtain exactly 23 students again who prefer to study with music playing if the study was repeated with a different group of 30 students.

> **Student 1**: "Because it's random, so the chance of getting 23 students that listen to music is possible, but not likely."

Student 2: "Because it's not going to be the same people in the study."

Student 3: "It's possible, but if random is random, you can't really decide which one is which."

The students in her class consistently stated their answers and contributed to the conversation, even if their thoughts were incomplete or not entirely accurate.

Robin and Karen have distinct techniques they used when questioning students during Lesson 4. Karen's techniques were a form of direct instruction, telling the students what to write and where. Robin invited students to give their own (and sometimes differing) opinions, encouraging them to write answers in their own words as long as the content and context of what they are saying was correct.

Pacing of the Lesson. **Karen**, as in Lesson 3, primarily controlled the pace of Lesson 4, mainly *departing* from the annotations and lesson plan. She used direct instruction which included providing step-by-step instructions throughout the majority of the activity. She read through all of the prompts and questions, controlling when students used the web applet or wrote down their answers on the activity. She had **an active and primary role in her students' learning**. During this lesson, her students had a very **limited role in their own learning** when she gave them a brief opportunity to explore with the web applet to learn about confidence levels, *adhering* to the lesson plan.

Additionally, during this lesson she asked students (after they shuffled the cards and recorded data) to write their re-sampled proportion on the whiteboard. This partially aligned with an annotation suggesting teachers use their students' re-sample proportions to construct a dot plot and explain how one of those dots was created. While Karen did

not create this plot, the fact that she had students write their different proportions down and then briefly discuss the variability between the sample proportions was a step in the right direction to motivate what students would see when using the web applet. This small shift in allowing students to take an active role in their own learning, in addition to using student data, was an improvement over Lesson 3, suggesting that Karen is **open to the idea of letting her students explore on their own, if time permits**.

Robin, more so than in Lesson 3, let her students control the pace of the lesson, *adhering* to the annotations and lesson plan. Students were working in groups throughout the lesson, **controlling the tempo of the lesson and engaging with the material**. Her students were taking an **active role in their own learning**. Robin was able to fully complete the lesson with time to spare. This extra time was spent giving students the sample homework questions (provided in the statistics unit) to work through, *departing* from the lesson plan and annotations, but in a way that was **reinforcing the learning objectives** and letting **students practice with the web applet and knowledge learned in Lesson 4**.

Karen and Robin controlled the pacing of the lesson in a way similar to their implementations of Lesson 3. Karen, constrained by time, controlled what she did or did not implement in the classroom. Robin, seemingly unconstrained by time, allowed the students to control the amount of time spent covering the material from Lesson 4. These differing techniques may be a direct result of the curriculum and type of schedule each teacher used. Karen was constrained by time in her Algebra II course using a traditional curriculum and Robin had more flexibility to cover more material in one day as she used

a block schedule (70-minute class periods) in Mathematics III under an integrated curriculum.

Content. **Karen** *adhered* to the content presented in the annotations and lesson plan in the same manner she did in Lesson 3. She followed the statistical content outlined in the lesson plan by using the activity and covered all of the learning objectives.

Robin *adhered* to the annotations and lesson plan in covering all of the content within these documents. Her content *diverged* when she **asked students to make connections to previous material**; namely Lesson 2 (*The Normal Distribution*). This occurred when students generated a re-sampling distribution; they were then asked to describe the shape, center, and spread of the distribution as they did when discussing the normal distribution in Lesson 2. She used this opportunity to further assess their understanding of the numbers used to describe the distribution. For example, students stated the spread was 0.073; Robin asked them to interpret what that number meant. She *departed* from the lesson plan when discussing the confidence intervals with students. Students were asked to calculate (by hand) a 95% confidence interval. She then asked them how they would go about creating a 68% confidence interval, referring students to think back to the empirical rule. Overall, Robin's departures revolved around making connections to previous material, a personal teaching goal of hers.

Overall Analysis of Karen and Robin's Implementations

Karen and Robin implemented the annotated lesson plans differently, adhering to and departing from the recommendations within (Tables 13 and 14). Generally, Karen

and her students used technology as a *productivity* tool, and her instructional actions mostly *departed* from the recommendations set forth in the annotations and lesson plan. Robin and her students used technology as a *cognitive* tool; here instructional actions primarily *adhered* to the recommendations in the annotations and lesson plan. As a result, they integrated technology at different levels, resulting in technology scores on opposite ends of the TPACK rubric.

I used the qualitative analysis to help explain *why* they implemented the annotated lesson plans differently. My aim was to investigate what characteristics of a given participant enabled her to use the annotated lesson plans in ways that guided the effective integration of technology when teaching statistical concepts. Part of my quantitative results suggested a participant's awareness of the use of inquiry was related to his or her integration of technology and corresponding technology scores. Through my qualitative analysis, I observed this characteristic to be more nuanced than simply having an awareness of its use. I found how a teacher's awareness of the use of inquiry is mediated by additional factors such as time constraints, curricular constraints, and a teacher's personal willingness to learn using inquiry.

Karen's use of technology as a productivity tool reduced the cognitive load of the activity. Her use of technology did not support inquiry-based learning; technology was used to quickly generate answers rather than to help students develop conceptual understanding. This use of technology may be a direct result of the training she acquired only involving use of graphing calculators, coupled with the emphasis on graphing calculators in AP Statistics.

Table 13. Summary of how Karen and Robin adhered to and departed from the annotations and lesson plan during their implementations of Lesson 3.

Instructional Action	**Karen**	**Robin**
Technology	**Adhered** *(in types used)***:** • Used web applet. **Departed** *(in purpose and use)***:** • Managed technological difficulties. • Relayed a misconception about how the web applet works. • Reduced the cognitive load of the web applet and activity. • Technology used for productivity and efficiency.	**Adhered** *(in types, purpose, and use)***:** • Used technology to develop students' conceptual understanding. • Helped students understand what technology was doing for them and why they were using it. • Technology used as a cognitive tool. **Departed** *(in use)***:** • Asked students to make connections to previous material.
Questioning	**Departed:** • Did not assess students' understanding.	**Adhered:** • Assessed students' understanding.
Information-Giving	**Adhered:** • Emphasized definitions of statistical terms, but in a manner that put primary focus on the definitions. **Departed:** • Provided answers. • Lowered the cognitive demand of the activity.	**Adhered:** • Emphasized definitions of statistical terms while keeping the focus on the primary learning objectives. **Departed:** • Provided additional information reflecting the the needs of her students.
Pacing of the Lesson	**Adhered:** • Managed class time more efficiently to help cover as much material as possible. **Departed:** • Teacher controlled the pace of the lesson. • Teacher had an active and primary role in students' learning.	**Adhered:** • Managed class time more efficiently to free up additional time to discuss statistical concepts and emphasize context. • Students controlled the pace of the lesson. • Students were engaged with the material. **Departed:** • Minimized the importance of a few questions.
Content	**Adhered:** • Covered material presented in the lesson plan. **Departed:** • Did not address all of the learning objectives.	**Adhered:** • Covered material presented in the lesson plan. • Addressed all of the learning objectives. **Departed:** • Covered additional material to have students make connections.

Table 14. Summary of how Karen and Robin adhered to and departed from the annotations and lesson plan during their implementations of Lesson 4.

Instructional Action	**Karen**	**Robin**
Technology	**Adhered** *(in use and types used)***:** • Used the web applet. • Used the "click one dot" feature in web applet, although this was stated *directly* in the activity. **Departed** *(in purpose)***:** • Relayed a misconception about how the web applet works. • Reduced the cognitive load of the web applet and activity.	**Adhered** *(in types, purpose, and use)***:** • Used technology to develop students' conceptual understanding. • Helped students understand what the technology was doing for them and why they were using it. • Used the web applet to its fullest potential
Questioning	**Adhered:** • Minimally assessed students' understanding.	**Adhered:** • Assessed students' understanding. • Encouraged students to take an active role in their learning. **Departed:** • Asked additional questions which reinforced the learning objectives.
Information-Giving	**Departed:** • Provided answers. • Lowered the cognitive demand of the activity. • Minimized the students' voice in the classroom.	**Adhered:** • using statistical terms and interpreting confidence intervals **Departed:** • Included additional questions. • Rearranged problems in the activity.
Pacing of the Lesson	**Adhered:** • Managed class time more efficiently to help cover as much material as possible. • Gave students a little time to explore and investigate. **Departed:** • Teacher controlled the pace of the lesson. • Teacher had an active role in students' learning.	**Adhered:** • Managed class time more efficiently to free up additional time to discuss statistical concepts and emphasize context. • Students controlled the pace of the lesson. • Students had an active role in their learning.
Content	**Adhered:** • Covered material presented in the lesson plan. • Addressed all of the learning objectives.	**Adhered:** • Covered material presented in the lesson plan. • Addressed all of the learning objectives. **Departed:** • Covered additional material to have students make connections.

The College Board for AP Statistics states how computers are an "essential tool for structured inquiry," yet, "under the conditions of a timed exam, students cannot be asked to perform the amount of computation that is needed for many statistical investigations" (College Board, 2010). As a result, the graphing calculator is the only tool available for students to use during the AP Statistics Exam. This influences the use of graphing calculators (in AP Statistics) to perform statistical analysis.

Throughout the entire unit, Karen demonstrated a clear understanding of how to use the graphing calculator for statistical computations. Her training and teaching experience in AP Statistics suggests she is familiar with the graphing calculator. This may explain her hesitation in using the web applet to teach statistical concepts, particularly when students will not be using computers for computations or analysis on standardized exams. Robin also teaches AP Statistics, but did not convey a hesitation to use the web applet as a learning tool. Through her surveys and interviews, she expressed that she has received training in using multiple types of technology as cognitive tools to support inquiry-based learning. Additionally, she has used simulation to teach statistical concepts prior to this experiment. These observations suggest she is familiar with more types of technology and has not been limited to using the graphing calculator.

Karen primarily used direct instruction to implement the annotated lesson plans; this was a result of the time constraints she felt were in place in Algebra II. She already had a large amount of content to cover in a limited amount of time, and the inclusion of a statistics unit added to this. She often provided students with answers and controlled the pace of the lessons by directing when students used the web applet and controlling how

far and deep they worked through the questions on the activity. Students were generally not provided the opportunity to explore and investigate with the web applet; this limited their ability to develop conceptual understanding of the statistical concepts and mechanisms underlying use of the computer-based technology.

I found it interesting that when Karen used the "click one dot" feature on the web applet and let students explore (both heavily emphasized in the annotations), questions in the activity *directly* stated to do so. I also noticed Karen had a hard copy of the activity and solutions present during each lesson, perhaps indicating a relative focus on the activity itself rather than the annotations. Accordingly, Karen may prefer direct instruction herself on how to effectively incorporate technology and inquiry in her class; in other words, perhaps she struggled with transforming knowledge stored in the annotations into her own classroom. When these "annotations" were directly integrated into a question on the activity, they were naturally incorporated into the lesson.

Robin primarily adhered to the annotations, and her use of them demonstrated that knowledge about teaching statistics with technology can indeed be stored in annotated lesson plans. For instance, she stated in an exit interview how she found the annotations, particularly the ones relating to the importance of one dot, as "very helpful." She gained knowledge in how to convey the meaning of one realization (i.e., what is one dot) from a sampling distribution to her students and translated this knowledge into her instruction. Furthermore, she continued to use the web applet in Lesson 5, in which no annotations or a technology-based activity were provided, to compare two treatments. She created a

technology-based activity for Lesson 5 that mirrored the ones provided in Lessons 3 and 4.

Yet, Robin's actions in and out of her class with the annotations demonstrated that, while knowledge about teaching statistics can be stored in annotations, it was not *directly* transferable. Robin conveyed use of a process of reading and questioning the annotations, working through the activity, and a need to think about how to best alter the activity and recommendations to incorporate them into her own classroom. Herein, it was vital to carefully consider the interaction of content, pedagogy, and technology. These actions allowed her to transform the stored knowledge into practice. Yet, she acknowledged that she spent time outside of school doing this, stating this is a process a "teacher has to go through if you are going to do a good job anyways."

With the adoption of the *CCSSM*, these statistics standards are now being introduced into Algebra II and Mathematics III courses. During this study, Karen and Robin demonstrated different perspectives about when statistics should be taught and how. Because they participated in this study and implemented the statistics unit in their Algebra II or Mathematics III course, they at some level, believe that statistics should be incorporated into these classes. Yet, they taught the content standards at differing proficiencies (which may result from their underlying perspectives of when students deeply need to understand the statistical ideas presented in this unit).

Karen's actions, particularly at the beginning and ending of Lesson 3, demonstrated hesitation of the statistics content being fully covered in Algebra II. She mentioned this was content her AP Statistics students see and failed to complete the

lesson in its entirety in her Algebra II class. She used approximately 20 minutes to implement Lesson 5 (*Comparing Two Treatments*), suggesting this is a content standard students will completely understand and do if they take AP Statistics. Through interviews, Robin expressed that she is trying to encourage more students to take her AP Statistics course. She expressed that statistics needs to be taught in all classes, stating, "This is [statistics] definitely part of what I have to teach. What we should teach." She also views it as more important than college algebra, and she has put forth effort to building a community within her class that views statistics as vital to creating citizens that can interact productively with society.

To some extent, Karen and Robin were both aware of using inquiry to teach and learn mathematics, although they were on opposite ends of the spectrum in regards to their awareness of the use of inquiry. Their differing experiences in professional development highlight these differing views. When asked about PD opportunities for teaching and learning statistics, Karen stated she has attended "only a couple of professional development workshops focusing on technology," whereas Robin stated the PD she has attended focused on technology and, more recently, teaching with "teams, tasks, and implementing the Teaching Practices from Principles to Actions."

Their differing experiences in PD have also influenced their personal views of using inquiry. For instance, Robin has found (through experience and PD) the benefits of using inquiry to teach and learn for herself and her students. She emphasized to her class and me that her role is to guide, rather than direct, students. In interviews, she described how she has been taking this journey towards "making sense of math" and views this as

one of the most important mathematical practices students learn in school. In her experience teaching, she has witnessed first-hand how rote memorization does not help students understand the material. She believes in the reform movement occurring in K-12 mathematics and statistics education; this is where students take an active role in their own learning, diverging from the traditional, formula-based approach. She recognizes the importance of "shift[ing] the way we are teaching" and, as a result, continues to work and learn ways to integrate this new approach in her teaching. These beliefs have positively influenced her use of the annotations and the use of technology to support inquiry-based learning throughout the statistics unit.

Karen and Robin's Exit Surveys

I asked each participant to complete an exit survey after they implemented all four statistics lessons. This survey contained the same inquiry and technology questions as the initial survey, though it also included one question about the integration of computer simulation in teaching and learning statistical concepts. I used this survey to assess the participants' post-experiment views of their awareness of the use of technology and inquiry in a mathematics classroom. Overall, these views were confirmed by the classroom observations, though in instances contradicted what I observed.

Awareness of the Use of Inquiry

While Robin's post-experiment views about the use of inquiry remained similar, I was surprised that Karen's views conveyed an increased awareness of the use of inquiry (in contrast to my observations of her class). In the exit survey, **Karen** did not mention

time constraints as a limitation on her use of inquiry. Rather, she implied that she uses inquiry and exploration tasks many times in her "statistics classroom." She first provided the following inquiry-based activity she uses which suggests a view of inquiry as not providing students any information.

> I start the year by allowing the students to design an experiment to see if the type/brand of putter you use in golf makes a significant difference in a golfer's ability to make the shot. I don't give them any information on the proper way to design an experiment, but I tell them to make it as legitimate as possible. Then, after we have learned about experiments and their design, we reflect on what they did and what they could have done to improve the study.

She further stated, "Problems dealing with probability and simulations are also good for letting the students explore and try to figure out on their own." This indicates that she views inquiry as including self-discovery (which was not emphasized in her initial survey). Her example of using simulation is interesting, as I did not observe her allow her students to use it to explore on their own. Instead, she directed, in all but one instance, how and when they should use the web applet.

All of the examples of inquiry tasks she provided involved statistics content, yet I did not observe her use much inquiry-based exploration in the statistics unit she implemented in her Algebra II course; I primarily observed her use direct instruction. This suggests she does use inquiry to teach statistics, but only in her AP Statistics course.

In her exit survey, Karen also expressed *why* inquiry-based explorations are important for teaching and learning mathematics; these views were not expressed in her initial survey. She stated:

> Both exploration and inquiry are extremely vital components when teaching and learning mathematics. The more opportunities that students are given to really think about the problems being presented and try to discover the solution

> themselves, the greater their chances of retaining that knowledge long into the future.

She elaborated on this by expressing the thought that teachers should allow students to solve problems themselves. Her responses indicated that she understands the value of using inquiry. Yet, I did not observe her give students the opportunity to "discover" solutions on their own. She would regularly give students answers to write down throughout the activity and demonstrate step-by-step how to use the web applet and complete the activity. Students were not engaged in the process of *doing* statistics themselves. These observations provide additional evidence that, while she may use inquiry in other classes, she does not use inquiry in Algebra II for the teaching and learning of statistical concepts.

Robin's post-experiment views remained consistent in her awareness of the use of inquiry. In her exit survey, she described the role exploration and inquiry have in teaching and learning mathematics:

> As I have progressed through my teaching career, it has become very evident that students do not truly learn from lecture, but rather from engaging and playing. The more opportunities I can give students to be given problems where they must engage in the standards of mathematical practice from the common core, the more likely they are to start questioning and making sense of the mathematics involved. Therefore, my classroom is structured to have students work regularly in teams and to minimize the amount of lecture time.

This aligned with what I observed in her classroom. She minimized the amount of direct instruction; this was particularly evident in Lesson 4, when she provided students more of an opportunity to explore and investigate with the web applet than she did in Lesson 3. Her experience teaching, along with her desire to help students understand what they are

doing, has led her to a particular method of teaching in which students explore with technology and are engaged with the material.

Awareness of the Use of Technology

Both Karen and Robin's post-experiment views about the use of technology did not vary much from their pre-experiment views. I found it particularly interesting that both teachers did not refer to the use of computer simulation (or the web applet) as a *productive* use of technology. This leads me to question how they would have taught these topics on their own; specifically, I wonder if they would have chosen to use a simulation-based approach.

Karen's responses to the technology questions continued to identify technology as a productivity tool. She viewed the use of graphing calculators as students' most productive use of technology, stating, "The most productive use of technology in the past 6 months is still the graphing calculators. They are used to input data and look at histograms and box-and-whisker plots, and derive equations that help make predictions." She also argued the role of technology is to "allow them [students] to compute and visualize their work more easily and quickly." This emphasis on efficiency aligned with her feeling of being time-constrained teaching Algebra II. Technology, on this end, appears to help alleviate this constraint through the quick calculation of answers. It may be the case that Karen is not aware of the use of technology to support inquiry-based learning, as this particular use of technology takes considerable time to implement.

Karen viewed technology as important to use in the classroom because it "surrounds us everyday." Her stance is that, because technology is "ingrained" in her

students' lives, it must be used in the classroom. As she stated, "students utilize technology even more than we do so it makes sense to incorporate technology into my classroom whenever possible." This reflects using technology for the sake of using it as opposed to using it to support and guide learning.

Robin's views remained consistent with those previously expressed. She demonstrated a high awareness of using technology to teach and learn mathematics. Her students were the primary users of technology, and it was used to help students develop conceptual understanding. Her examples of the most productive uses of technology included technology as a cognitive tool. She stated how "Students can play with math using technology individually to see how a concept manifests itself or to be able to move from reasoning quantitatively to reasoning abstractly." Her emphasis on "play" indicates that her students really do explore with technology.

Robin identified the ability to "see different results quickly" as one benefit of technology. During my observations, Robin had students use the extra time afforded by technology to "share their thinking with others and hear how others may be thinking about what they see differently." She identified the need for students "to be able to use the technology to explore the mathematics as well as give them results without their need to have a great amount of skill in the procedural aspects of mathematics, such as algebra." I interpreted this as a belief that *all* of her students in *all* of her classes can use technology to their benefit and help them (better) understand mathematical and statistical concepts.

Overall Summary

The goal of this research study was to investigate the extent to which knowledge about teaching statistics with technology is passed along to in-service teachers via specially annotated lesson plans. The process of helping teachers effectively use technology in the instruction of statistics is not straightforward. Teachers vary in many characteristics, adding to the complexity of improving methods of instruction that will help all teachers.

The quantitative analysis used data from five volunteer participants who were randomly assigned to receive the annotated or non-annotated statistics unit; consequently, conclusions suggested by statistical analysis can be inferred to this sample of high school teachers. Yet, this is not to say that recommendations cannot be made to teachers, researchers, and educators about the use of annotated lesson plans. The qualitative analysis used data from two selected participants and helped convey why the two teachers implemented the annotated lesson plans differently, describing in detail how they adhered to and departed from the annotated lesson plans; I highlighted reasoning, such as time constraints, technological difficulties, teaching methods, and so on, for doing so.

The quantitative results suggested no difference in technology scores between those who received the annotated statistics unit and those who did not; there was also a wide variation in how teachers integrated technology during the four lessons. This group of five teachers generally taught Lesson 2 in a similar manner, using technology as a productivity tool rather than a cognitive tool. In both Lessons 3 and 4, I observed most of

the teachers integrate technology *more* effectively than they did in Lessons 2 and 5. In fact, due to the nature of the technology-based activities, all of these teachers used a computer simulation (in some manner) to teach the statistical content. In all but one instance, these teachers and their students employed a dynamic use of technology during Lessons 3 and 4 by conducting a simulation to generate a re-sampling or re-randomization distribution. I observed teachers implement Lesson 5 in a variety of ways, some with and without the use of cognitive technologies. Two participants demonstrated a positive change in their technology scores from Lesson 2 to Lesson 5; one had received an annotated statistics unit and the other had not.

I also examined the relationship between the three teacher characteristics – awareness of the use of inquiry, awareness of the use of technology, and statistical content knowledge – and the participants' integrations of technology. The data suggested that a participant's inquiry ranking may be related to his or her total technology scores; those with a higher awareness of the use of inquiry integrated technology more effectively than those with a lower awareness. I did not observe the same relationship between my participants' technology rankings or statistical content knowledge scores and their total technology scores. In fact, of the two teachers from my qualitative analysis, the teacher with the higher statistical content knowledge score demonstrated more misconceptions about sampling distributions than the other teacher with a lower statistical content knowledge score. This suggests the assessment I used was not fully representative of a teachers' statistical content knowledge, particularly with sampling distributions.

Lastly, I explored the average time the five participants spent implementing the four statistics lessons versus their average total technology score. Only one teacher used more than 100 minutes to implement each of the lessons; this teacher displayed the most effective integration of technology as she used technology to support inquiry-based learning. Three of the remaining participants spent, on average, one 50-minute class period, on each of the lessons. Their average total technology score was representative of the lower levels of the TPACK rubric. This suggests that teachers need considerable time to effectively integrate technology in their teaching of a statistics unit.

The participants I chose for the qualitative analysis had different teaching strategies and were on opposite ends of the spectrum regarding their awareness of the use of inquiry; one relied primarily on direct instruction, the other, inquiry exploration tasks. Throughout both Lessons 3 and 4, I continually observed the former depart from the annotated lesson plans. The latter teacher adhered to them, resulting in a more effective integration of technology. Here, I was also able to observe *how* and *when* teachers implemented inquiry-based explorations (as opposed to how they viewed the use of inquiry in the quantitative results). The teacher who valued inquiry in learning and teaching facilitated her students' use of inquiry for the learning of statistical concepts. The teacher who expressed time constraints did not use inquiry in her implementation of the statistics unit.

Annotations are capable of storing knowledge about teaching statistics with technology; however, the applicability of a teacher transferring this knowledge into their teaching methods and classroom was influenced by many factors. These included a

teacher's awareness of the use of inquiry, enactment of inquiry in the classroom, time constraints, experience with technology and inquiry, view of incorporating statistics content standards in Algebra II or Mathematics III, and his or her values as a teacher. In other words, the recommendations in the annotations were not directly transferable. The teacher who recognized and valued the change in approach (from traditional to data-oriented) and used inquiry-based explorations in Mathematics III demonstrated an ability to transfer the stored knowledge, albeit only after considering how the annotated lesson plan and technology-based activity fit into her own classroom and preferred method of teaching. Clearly, this process took time to think about and to implement; it was not something that happened instantaneously. The teacher who was constrained by time demonstrated hesitation about using computers over calculators and teaching the statistics content standards in Algebra II. She relied on direct instruction to cover as much material as possible, and was less likely to implement recommendations from the annotations unless they were directly stated in the technology-based activity.

CHAPTER 5

CONCLUSIONS

Introduction

How can researchers and teacher educators help foster and develop in-service high school teachers' ability to effectively use technology when teaching statistics? is the question that has motivated my research. With the adoption of national and state content standards, high school teachers are now expected to teach statistical concepts in mathematics courses such as Algebra II and Mathematics III and are encouraged by many national organizations to implement a data-oriented approach in their instruction. Yet, statistics is still a relatively new subject for many in-service high school teachers (Franklin et al., 2007). Few have been taught using this data-oriented approach they are now being asked to teach (Franklin et al., 2015). Hence, there is a great need to assist in-service high school teachers by, for instance, developing and disseminating materials to support the instruction of statistics with technology.

In this study, I investigated (via a completely randomized block design) whether specially annotated lesson plans were a means to store knowledge of teaching statistics with technology and guide in-service high school teachers' effective use of technology. I quantitatively analyzed five high school teachers' integration of technology with or without annotated lesson plans in the teaching of statistics content standards in an Algebra II or Mathematics III course. A qualitative analysis of data from two of these teachers provided substance and context to the quantitative results. In this chapter, I will

review the key results, discuss implications for teacher educators and researchers, provide ideas for future research, and identify some limitations of the study.

Summary of Key Results

The graphical analysis of the quantitative data suggested no clear differences in technology scores between those who received and those who did not receive the annotated statistics unit. I observed variation among participants' technology scores indicating some differences in how they integrated technology during the lessons. However, all participants, regardless of received unit, integrated technology more effectively when they were provided with a technology-based activity employing computer simulation. I also observed the disparity in participants' awareness of the use of inquiry and their enactment of inquiry-based explorations and how these related to their integration of technology throughout the statistics unit. Participants who conveyed a higher awareness of using inquiry in a mathematics classroom and those who demonstrated using inquiry in their implementations tended to integrate technology more effectively than those with a lower awareness and less frequent (or non-existent) use of inquiry. The qualitative analysis revealed many factors (such as time constraints, curricular constraints, and a teacher's personal willingness to teach and learn using inquiry) that were related to a participant's decision to use technology and inquiry-based explorations to teach statistics in her Algebra II or Mathematics III classroom.

Type of Statistics Curriculum

Many national organizations recommend statistics content standards to be taught in a setting where statistical thinking and conceptual understanding are emphasized and inquiry-based learning is used to explore data, making effective use of technology. There were two different types of curriculum used in this study – one that is conducive to the goals of these organizations and another that is not. The type of curriculum participants used appeared to motivate the amount of time they spent implementing the unit, impacting their instructional decisions and ultimately their integration of technology. As a result, the different types of curricula used by participants may have contributed to the differing technology scores received.

Robin, the only participant who integrated technology at the highest level of the TPACK rubric, taught Mathematics III under an integrated curriculum, in which multiple topics are combined into one class. Her curriculum already had the statistics content standards naturally integrated, and she was able to use more than a week for implementation. In fact, she used approximately two 70-minute class periods to teach each of the four statistics lessons. Many of the other participants taught Algebra II under a traditional curriculum, where these statistics content standards have not been incorporated. Furthermore, there is a large amount of existing curricular material to cover in Algebra II. Inclusion of this statistics unit in Algebra II, therefore, appeared to be haphazardly "forced in" with direct instruction by many of the participants.

Use of the integrated curriculum has some immediate benefits for teaching statistics with an inquiry-based approach; there is more freedom for a teacher to jump

from topic to topic, establishing connections as multiple topics are taught. Furthermore, the textbook often used within this curriculum prompts students to justify their answers and work collaboratively. In such environments, there is often an abundance of exploration and a corresponding lack of rote memorization of formulas. These characteristics describe a classroom and curriculum well-suited for statistics instruction at the levels recommended by the *Pre-K-12 GAISE Report*, *CCSSM*, and the *SET Report*. Traditional lecture, on the other hand, is neither the desired nor optimal mode of instruction for these standards given the goals prescribed by the aforementioned organizations.

Implications

The results from this study have several implications for both teacher educators and researchers working to improve the instruction of statistics with technology.

Implications for Teacher Educators

First and foremost, it is imperative for schools and teachers to develop a culture in which statistics is viewed as equally important as mathematics. At the high school level, this view is not fully expressed if there is a belief that only students who take an AP statistics course should be exposed to statistics or if there is a minimal amount of time allotted for its implementation. The statistics content standards are meant to be taught *throughout* all high school classes; high school teachers will be more inclined to teach these content standards if they believe statistics is important for students to learn and have the support of their schools to integrate these standards into their curriculum.

The way statistics is recommended to be taught (using a data and simulation-based approach) requires time to implement well. The statistics content standards are meant to be introduced more intuitively than how they are in an AP Statistics course (Franklin, 2014; Franklin et al., 2015). Therefore, teachers need to have an appropriate amount of time available to fully implement these standards. The only teacher to integrate technology at the highest levels of the TPACK rubric used on average, more than one class period to implement each lesson. This time was already built into her curriculum, indirectly providing her support to implement these standards as intended. The six teachers who received but did not implement a statistics unit were unable to find a time to fit it into their Algebra II courses during the 2015-2016 academic year. Many indicated an urgency to first cover content that was going to be assessed by standardized testing.

As teachers are exposed to the value of teaching statistics and are encouraged to integrate the statistics content standards into their classrooms, they need training in teaching statistics with technology and at the levels suggested by the *Pre-K-12 GAISE Report* and the *CCSSM*. These reports recommend, for instance, that teachers and students use appropriate technological tools to perform simulations to explore statistical concepts, developing conceptual understanding in the process. Yet, in-service teachers are often unfamiliar with the use of computer simulation for this purpose, despite its advantage in helping teachers become familiar with this approach. One of the findings from this study was that technology-based activities helped *all* teachers integrate technology more effectively through the use of computer simulation; this was evident in the higher technology scores for Lessons 3 and 4 compared to Lessons 2 and 5. The

participants in this study, even those not well-acquainted with simulation, had positive things to say about its use for teaching statistical concepts and implied they would continue to use simulation to teach statistics in the future.

The use of simulation is best taught with an inquiry-based approach to learning, as students must understand what technology is doing for them in a data-oriented context, and exploring allows them to see this more readily. The statistics unit naturally fit into an inquiry-oriented classroom and the results of this study highlighted how a participant's awareness and use of inquiry related to their integration of technology with or without annotations. Those with a higher awareness of using inquiry integrated technology more effectively. Moreso, the participant who actively *used* inquiry in her classroom (Robin) received technology scores aligned with the higher levels of the TPACK rubric. Like Robin, teachers who understand why to use inquiry in their classrooms *and* enact inquiry will likely possess skills to effectively integrate technology and ask students questions that further their exploration and hence understanding of the material. Therefore, teacher educators should help teachers recognize the advantages of using inquiry to learn statistics and be trained in implementing inquiry-based learning in their classrooms.

Based on the classroom observations, many of the teachers appeared most confident teaching Lesson 2, and they all taught it similarly. Its content revolved around the normal distribution, for which there are many teaching resources available with or without technology. Conversely, there was more variation in how the participants implemented Lesson 5, and there are not as many resources for this topic as there are for Lesson 2. With the exception of the second lesson, this was the only other lesson that did

not have a corresponding technology-based activity; yet, many participants asked to be provided with one for Lesson 5. Due to my research design, it was not possible to provide this. Those that did not create an activity for this lesson primarily used direct instruction and lectured during Lesson 5, resulting in a drastic decrease of technology scores from Lesson 4 to 5. It is important to note that teachers often have limited time to develop activities on their own. Thus, to support the implementation of a data-oriented approach in teaching statistics, the development of technology-based activities has great potential in aiding in-service teachers of statistics.

Statistics is not an easy subject to teach and make sense of; another way to support statistics instruction with technology is through the development of *annotated* lesson plans that encourage the use of these technology-based activities to support inquiry-based learning. For example, pointed questioning by the teacher can help reinforce statistical concepts developed through simulation, something teacher educators may help the greater population of teachers realize and prepare for via annotations. The results of this study indicated that annotated lesson plans are a useful vessel for knowledge about teaching statistics effectively with technology that can then be implemented into practice by those willing. For example, Robin demonstrated reading and applying the annotations in her own instruction, but she already had a method of teaching that was conducive to inquiry-based learning with technology. She recognized the importance of using inquiry and exploration for the teaching and learning of statistics. As a result, she more readily translated and transformed the knowledge stored in the annotations.

Lastly, Karen relied heavily on the use of graphing calculators, which was reasonable due to her familiarity with it induced by extensive use in AP Statistics. Her reluctance to use the web applet instead of the graphing calculator is likely to be representative of many current AP Statistics teachers who are being asked to teach statistics in non-AP Statistics courses. As statistics education shifts to using computers and other types of technology (such as web applets) to perform computations and statistical analysis, teacher educators should consider the impact of the current teachers who primarily use graphing calculators. Teacher educators should be considerate of this hesitation and focus professional development workshops on mitigating it and encouraging teachers to switch to computers.

Implications for Researchers

The attrition rate of participants in this study was high. To improve this situation, researchers can provide more incentives. Alternatively, it may be beneficial for researchers to develop a personal relationship with participants. I was not able to visit participants' classrooms to observe them teach; all of the observations occurred remotely, via FaceTime or video-recordings. This excluded a personal connection between the researcher and participant, possibly raising the attrition rate of my participants.

Asking participants to record themselves on video proved more challenging than I initially anticipated. Some of the videos were incomplete or poorly recorded. By being in the classroom, I would have made complete observations and been able to view interactions between the teacher and students. Finally, some participants were not

comfortable (or were not able to receive permission from administration) being video-taped and opted out of the study during the recruitment phase.

I only communicated with teachers (as opposed to administration) during the recruitment phase. Reaching out to administration may have helped in obtaining more teachers to participate. The final group of participants was naturally inclined to implement a statistics unit, yet many teachers are hesitant about doing so. If they had the encouragement of their administration, they may have been more likely to participate.

Often, one of the steps in designing a study involves the calculation of a sample size. I did not have any pilot data to use to make an informed calculation of sample size in my design. However, the data from this research can provide researchers with pilot data to use for future sample size calculations. Researchers may use my data to obtain a very conservative estimate of the residual standard error by calculating the standard deviation of the appropriate technology scores. However, the sample size calculations should be consistent with the research instruments chosen for use and the intended type of statistical analysis to be employed. As a result, researchers will need to judge the applicability of my results to their sample size calculations. Researchers should also use caution, however, with the sample size they calculate. If using a conservative estimate of the residual standard error, the sample size calculation will likely result in a sample size that may be too large to feasibly attain.

In my study design, I blocked participants into pairs prior to random assignment of the statistics units. There are potential disadvantages to this; namely, there is no replication within each block and treatment combination. Replication refers to assigning

each treatment condition to more than one participant within a block. By pairing participants, only *one* participant in each block receives an annotated or non-annotated statistics unit. However, replication is sometimes difficult to achieve with a small number of participants or prohibitive cost in treatment. For this study, because of the difficulty in recruiting participants and the small number who volunteered to participate, I made a decision to forego sorting participants into less homogeneous blocks (for replication). This, however, favorably allowed me to control for the negative effects of confounding variables and extra sources of background variability (such as the participants' awareness of the use of inquiry, awareness of the use of technology, and statistical content knowledge) in my design. Therefore, it would have been easier to detect any treatment effects had I retained all my participants within each block.

The blocks I created were based on a teacher's awareness of the use of inquiry, awareness of the use of technology, and statistical content knowledge. Through the quantitative and qualitative analysis, I observed that a teacher's awareness of the use of inquiry *and* his or her enactment of inquiry related to the manner in which technology was implemented. I assessed the participants' awareness of the use of inquiry with two open-ended inquiry questions, and their initial responses tended to correspond closely with my observations of the teachers' use of inquiry in their Algebra II or Mathematics III classrooms. Therefore, I recommend researchers continue to assess a teacher's awareness of the use of inquiry at the beginning of the study and focus on using those responses to block participants in future studies. This will allow for comparison between

annotated and non-annotated groups participants with similar levels of awareness of the use of inquiry.

Researchers may also reconsider the type of model to employ or change the nature of the response variable through a different TPACK rubric. The rubric I used provided quantitative data in the form of ordered categorical values. An ordered categorical logistic regression model, for instance, may be more appropriate to model the data collected. However, with any model comes assumptions. The commonly used ordered categorical regression models carries a proportional odds assumption, which assumes the relationship between any two categories is the same. For example, in the context of the TPACK rubric, one would need to assume the relationship from the *Recognizing (1)* level to the *Accepting (2)* level is the same as the relationship from the *Adapting (3)* level to the *Exploring/Advancing (4)* level. Future researchers using a TPACK rubric ought to consider the assumptions and trade-offs when deciding on a model and rubric.

For the purpose of this study, I adapted a rubric used to assess a teacher's TPACK in mathematics; further observations of statistics instruction may aid in the development and verification of a rubric for a teacher's TPACK in *statistics*. For instance, graphing calculators are often a staple of statistics classrooms, and while they are used primarily as a productivity tool (corresponding to the two lowest levels of the rubric), it was difficult to distinguish between these with respect to calculator use alone. Furthermore, when using the rubric, I chose to consider both performance indicators together when assigning

scores. This simplified the evaluation of each teacher's integration of technology at the cost of increased discretization of the rubric.

Improving Teaching versus Improving Teachers

The use of annotated lesson plans was motivated by Hiebert and Morris (2012) who proposed an alternative approach to improve teaching. They recommend that, to improve teaching, we must focus on the instructional methods that are implemented in the classroom rather than the characteristics of teachers. Lampert (2012) and Lewis, Perry, Friedkin, and Roth (2012) add to this conversation, responding to Hiebert and Morris' article (2012). Lampert (2012) argues that improving teaching necessarily implies that the characteristics of teachers are improved, and that teachers need to be committed to teaching well. Lewis et al. (2012) argue that annotated lesson plans need to be accompanied by opportunities for teachers to see and try these new approaches in practice. The results from my study add to this conversation about using annotated lesson plans to improve teaching.

I observed firsthand how teachers' commitment and will to teach statistics in Algebra II was related to their implementation of the statistics lessons. One participant, who had never taught statistics, was committed to implementing the statistics unit in its entirety and did so despite a lack of confidence in her instruction. Conversely, the participants who did not implement the statistics unit did not have a strong commitment to implement the statistics unit. Part of this hesitation was prioritized by the need to teach content covered on standardized testing, delaying their planned implementation until it

was not possible. A teacher's will may also be motivated by that of their district. If a school embodies a culture viewing statistics as non-essential, its teachers will be less motivated to implement statistics content standards.

Teachers need to be motivated to try new approaches and work on improving their teaching. Lampert (2012) discusses how the cognitive demand of the materials is lowered when students become "resistant to the different and more difficult challenges of unfamiliar tasks." I observed this in Karen's class during implementation of the statistics lessons, as her students were primarily used to direct instruction and the use of calculators, procedures, and formulas opposed to computer simulation. Teaching and learning statistics at the levels suggested by the *Pre-K-12 GAISE Report* and the *CCSSM* is ultimately more difficult and challenging than the traditional formula-based approach.

Future Research

This research demonstrated the efficacy of annotated lesson plans as a store of knowledge about teaching statistics with technology. With the information I collected during the classroom observations, I plan to revise the annotations, limiting the amount of variation observed in my participants' implementation of them. I will update the plans with useful knowledge (particularly about simulations) that can guide teachers' implementation of statistics using technology. For example, I plan to incorporate more questions teachers could ask their students about the use of a simulation and how it creates a sampling distribution. Secondly, I plan to begin disseminating the annotated

lesson plans and continue research about how and why teachers do or do make use of the annotations.

For this research, I was the primary individual engaged in directly creating these annotations. I only studied what occurred as a teacher implemented an annotated lesson plan, not what they would change after implementation. The intended cycle of annotated lesson plans is for teachers take part of the process of creating, implementing, and revising them with the most current and relevant knowledge. It would, therefore, be beneficial to work with a group of teachers, perhaps from the same mathematics department, collaboratively studying and improving the annotations together as they are being implemented. I wonder if it is possible for teachers to acquire new knowledge of content, pedagogy, and technology and develop a deeper understanding of how to effectively integrate technology when they engage in a form of "lesson study" with the annotated lesson plans?

Future research in this field could also target different populations of teachers to gather more knowledge of how and when annotated lesson plans improve the instruction of statistics with technology. The five participants who volunteered in this study clearly demonstrated an interest in teaching statistics content standards in Algebra II or Mathematics III. There are, however, high school teachers who are not necessarily confident in teaching statistics; this is a group that is desperately in need of methods to improve instruction. It would be valuable to target this population in a future study. Additionally, I observed how a participant's awareness of the use of inquiry *and* their enactment of inquiry in the classroom were related to their integration of technology, and

I am particularly interested in learning more about how inquiry is related to a teacher's integration of technology. For this purpose, it would be beneficial for me to try and retain *pairs* of participants so I can make effective comparisons between technology scores of those who received the annotated statistics unit and those who did not after accounting for a teacher's awareness of the use of inquiry. I plan to begin by asking the six participants who were not able to implement the statistics unit if they are still interested in doing so.

Teaching statistics differs from teaching mathematics (Franklin et al., 2007, 2015). Hence, a teacher's TPACK in statistics is different from his or her TPACK in mathematics. Researchers have examined and proposed different frameworks for technological pedagogical content knowledge in statistics, or TPSK (e.g., Lee and Hollebrand, 2011; Lesser and Groth, 2008). The TPACK framework appears to display content knowledge, pedagogical knowledge, and technological knowledge with the same weight and, thus, relationship to a teacher's overall TPACK. However, the results of this study suggest a teacher's awareness of inquiry (one facet of his or her pedagogical content knowledge) has a more profound relationship on TPACK in statistics than content or technological knowledge. This research may aid in investigating how TPACK differs between statistics and mathematics and what influence, if any, this exerts on assessment of TPACK in statistics. Therefore, this research may be helpful for research surrounding the TPSK framework and the development of rubrics to measure TPSK.

Limitations

The purpose of this study was to investigate annotated lesson plans as a means to store knowledge and guide effective integration of technology in teaching statistics. This study adds to the literature on annotated lesson plans, developing technology-based activities for statistics instruction, what affects a teacher's enactment of inquiry curriculum, and TPACK in statistics and mathematics. It is also valuable for teacher educators who are training and providing professional development for in-service K-12 teachers who are now being asked to teach statistics content standards in a way they themselves may not have learned; this involves the use of technology (and simulation) to support inquiry-based learning. With that being said, I note some potential limitations of this research study.

Teaching is a multifaceted profession, and the characteristics, beliefs, and classroom structures of teachers vary tremendously, with no two being identical. This adds marked difficulty for one to apply results from research to all types of teachers and classrooms. The five participants in this study are not representative of the general population of high school teachers in the U.S. These five teachers were the few (out of hundreds emailed) who had access to class sets of tablets or computers and chose to participate. While one of the participants explicitly told me she had a personal inclination to participate, she also expressed a professional desire to participate; she indicated her interest in teaching statistics and trying new techniques and materials in her classroom. This interest in teaching statistics in Algebra II and Mathematics III was common among

each of the five participants. Each viewed statistics as an integral part of the K-12 mathematics curriculum, even if their experience in teaching statistics varied greatly.

Next, I did not do a post-interview with Karen to gather additional insight about her instructional decisions and views of using inquiry. The interview with Robin provided insight into the characteristics and beliefs of a teacher who used and applied the knowledge from the annotations into her instruction of statistical concepts. Her interview helped identify the contexts in which annotations may be useful for guiding the effective integration of technology. Obtaining an interview with Karen may have been useful to help identify the contexts in which annotations may *not* be useful. This information would be invaluable in making revisions to the annotations so they are useful for all teachers.

Concluding Statements

Annotated lesson plans are a means to store knowledge about teaching statistics with technology, and it is possible for this knowledge to be translated by teachers and transferred into practice. The degree of transferability may be related to a teacher's awareness of the use of inquiry and their enactment of inquiry in the classroom. Using a data-oriented approach to teach statistics with technology is naturally related to the use of inquiry to develop students' conceptual understanding. Many of the questions posed in the annotations and technology-based activities do not have simple answers, but they force students to interact with the context of the problem and what technology is doing for them "behind the scenes" (for instance, how the re-sampling distribution is created).

The coverage of the statistical concepts within the statistics unit takes time, and hence it is not well-suited for brief direct instruction with limited student engagement. Teachers who are aware of the benefits of using technology and make use of inquiry-based learning in all of their classes will more likely apply the knowledge in the annotations into practice, as these annotations naturally align with the use of technology in inquiry-based learning.

With five total participants, I did not have enough observations to detect a difference (if one existed) in mean technology scores between the annotated and non-annotated group. However, the unit I developed proved to be very valuable for all teachers involved in the study. The technology-based activities, specifically, allowed teachers to integrate technology more effectively than without. With these activities, teachers at all levels of experience teaching statistics were using computer simulation to implement a data-oriented approach to the instruction of statistics.

It is clear that to produce statistically literate students, K-12 teachers need to incorporate statistics content standards into their mathematics classes, ideally with the use of technology-based activities. Teachers, however, often lack time or the relevant knowledge to develop these activities. It is thus important for teacher educators and researchers to work collaboratively, providing in-service teachers with instructional methods (such as annotated lesson plans containing technology-based activities) that not only guide their teaching of statistics, but guide their effective use of technology.

Lastly, the integration of technology appeared to be most related to participants' awareness of the use of inquiry and their enactment of inquiry in the classroom. Focusing

on these attributes in a professional development workshop may greatly improve teachers' integration of technology in teaching for a data-oriented statistics curriculum. Ultimately, different teaching methods are best suited for different subjects. The teaching of statistics differs from the teaching of mathematics, and the use of inquiry-based learning is well suited for teaching statistics from a data-oriented perspective. Students need to be pushed to reason statistically; this is accomplished much more easily when a teacher is given time to employ inquiry-based exploration in their classes.

REFERENCES CITED

Abbitt, J. (2011). Measuring technological pedagogical content knowledge in preservice teacher education: A review of current methods and instruments. *Journal of Research on Technology in Education*, *43*, 281-300.

Agresti, A. (2007). *An introduction to categorical data analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

American Statistical Association-National Council of Teachers of Mathematics. (2013). *Preparing Pre-K-12 Teachers of Statistics*. A Joint Position Statement of the American Statistical Association (ASA) and the National Council of Teachers of Mathematics (NCTM).

Angeli, C., & Valanides, N. (2005). Preservice elementary teachers as information and communication technology designers: An instructional systems design model based on an expanded view of pedagogical content knowledge. *Journal of Computer Assisted Learning*, *21*, 292 - 302.

Association of Mathematics Teacher Educators. (2015). *Position of the Association of Mathematics Teacher Educators on Technology.* Raleigh, NC: Author.

Bargagliotti, A. (2014). Statistics: The new "it" common-core subject. *Education Week*, *33*(19), 20-21.

Bargagliotti, A., & Franklin, C. (2015). The statistical education of teachers: Preparing teachers to teach statistics. *Chance*, *28*(3), 19-27.

Bargagliotti, A., Jacobbe, T., & Webb, D. (2014). A commentary on elementary teacher preparation to teach statistics. *Issues in the Undergraduate Mathematics Preparation of School Teachers, 4.*

Barrow, L. (2006). A brief history of inquiry: From Dewey to standards. *Journal of Science Teacher Education*, *17*, 265-278.

Batanero, C., Burrill, G., & Reading, C. (2011). Overview: Challenges for teaching statistics in school mathematics and preparing mathematics teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICMI/IASE study.*

Batanero, C., Burrill, G., Reading, C., & Rossman, A. (2008). Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. *Proceedings of the ICMI Study 18 and IASE 2008 Round Table Conference.* Monterey, Mexico, International Commission on Mathematical Instruction and International Association for Statistical Education.

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenges of developing statistical literacy, reasoning and thinking*. Dordrecht: Kluwer Academic.

Brantley-Dias, L., & Ertmer, P. (2013). Goldilocks and TPACK: Is the construct "just right?" *Journal of Research on Technology in Education*, *46*, 103-128.

Carver, R., Everson, M., Gabrosek, J., Rowell, G., Horton, N., Lock, R., … , Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE): College Report.* Alexandria, VA: American Statistical Association.

Chance, B. (2002). Components of statistical thinking and implication for instruction and assessment. *Journal of Statistics Education*, *10*(3).

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).

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

Conference Board of the Mathematical Sciences (2001). CBMS. *Mathematical Education of Teachers I.* Providence RI and Washington DC: American Mathematical Society and Mathematical Association of America.

Conference Board of the Mathematical Sciences (2012). CBMS. *Mathematical Education of Teachers II.* Providence RI and Washington, DC: American Mathematical Society and Mathematical Association of America.

College Board. (2010). *Statistics Course Description*. The College Board. Retrieved from http://www.collegeboard.org.

Creswell, J. (2013). *Qualitative inquiry & research design: Choosing among five approaches*. Los Angeles, CA: SAGE Publications.

Davis, R. (1998). Prayer can lower blood pressure. *USA Today*, August 11.

delMas, R. (2002). Statistical literacy, reasoning, and learning: A commentary. *Journal of Statistics Education*, *10*(3).

delMas, R. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenges of developing statistical literacy, reasoning and thinking*. Dordrecht: Kluwer Academic.

Dewey, J. (1910). Science as subject-matter and as method. *Science*, *31,* 121–127.

Ellis, A., Bieda, K., & Knuth, E. (2012). *Developing essential understanding of proof and proving.* Reston, VA: National Council of Teachers of Mathematics.

Franklin, C. (2000). Are our teachers prepared to provide instruction in statistics at the K-12 level? *NCTM Mathematics Education Dialogues*, *37*(3).

Franklin, C. (2013). Common core state standards and the future of teacher preparation in statistics. *The Mathematics Educator*, *22*(2), 3-10.

Franklin, C. (2014). The statistical education of teachers (SET): An American Statistical Association policy document. In K. Makar, B. de Sousa, & R. Gould (Eds.), Sustainability in statistics education. *Proceedings of the Ninth International Conference on Teaching Statistics* (ICOTS9, July, 2014), *Flagstaff, AZ.* Voorburg, The Netherlands: International Statistical Institute.

Franklin, C., & Kader, G. (2010). *Models of teacher preparation designed around the GAISE framework.* Paper presented at the 8th International Conference on Teaching Statistics, Ljubljana, Slovenia.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A pre-K-12 Curriculum Framework*. Alexandria, VA: American Statistical Association.

Franklin, C., Bargagliotti, A., Case, C., Kader, G., Scheaffer, R., & Spangler, D. (2015). *Statistical Education of Teachers (SET) Report*. Alexandria, VA: American Statistical Association.

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, *10*(3).

Garfield, J., Aliaga, M., Cobb, G., Cuff, C., Gould, R., Lock, R., …, Witmer, J. (2005). *Guidelines for Assessment and Instruction in Statistics Education (GAISE): College Report*. Alexandria, VA: American Statistical Association.

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*. Springer.

Garfield, J., delMas, R., & Chance, B. (2003). *Web-based assessment resource tools for improving statistical thinking.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Garfield, J., delMas, R., & Chance, B. (2006). The Web-based ARTIST: Assessment Resource Tools for Improving Statistical Thinking Project.

Garfield, J., delMas, R., & Zieffler, A. (2008). *The CATALST Project: Change Agents for Teaching and Learning Statistics,* National Science Foundation, DUE 0814433.

Groth, R. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, *38*(5), 427-437.

Groth, R., & Bargagliotti, A. (2012). GAISEing into the common core of statistics. *Mathematics Teaching in the Middle School*, *18*(1), 38-45.

Hiebert, J., & Morris, A. (2009). Building a knowledge base for teacher education: An experience in K-8 mathematics teacher preparation. *The Elementary School Journal*, *109*(5), 475-490.

Hiebert, J., & Morris, A. (2012). Teaching, rather than teachers, as a path toward improving classroom instruction. *Journal of Teacher Education*, *63*(2), 92-102.

Hollebrands, K., & Zbiek, R. (2004). Teaching mathematics with technology: An evidence-based road map for the journey. In R. Rubenstein & G. Bright (Eds.), *Perspectives on the teaching of mathematics, 2004 yearbook of the National Council of Teachers of Mathematics* (p. 259-270). Reston, VA: National Council of Teacher of Mathematics.

Hu, Y., & Hu, F. (2012). Balancing treatment allocation over continuous covariates: A new imbalance measure for minimization. *Journal of Probability and Statistics*, *2012*.

Kader, G., & Perry, M. (2006). A framework for teaching statistics within the K-12 mathematics curriculum. *Proceedings from the ICOTS-7 Conference.*

Koehler, M., & Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research*, *32*(2), 131-152.

Koehler, M., Shin, T., & Mishra, P. (2011). How do we measure TPACK? Let me count the ways. In R. Ronau, C. Rakes, & M. Niess (Eds.), *Educational technology, teacher knowledge, and classroom impact: A research handbook on framework and approaches* (p. 16 - 31). Hershey, PA: IGI Global.

Lampert, M. (2012). Improving teaching and teachers: A "generative dance"? *Journal of Teacher Eduation*, *63*(5), 361-367.

Larson, R., Boswell, L., Kanold, T. D., & Stiff, L. (2012). *Holt McDougal Larson Algebra 2*. Orlando, FL: Holt McDougal – Houghton Mifflin Harcourt.

Lazar, N., & Franklin, C. (2015). The big picture: Preparing students for a data-centric world. *Chance*, *28*(4), 43-45.

Leavy, A., Hannigan, A., & Fitzmaurice, O. (2013). If you're doubting yourself then, what's the fun in that? An exploration of why prospective secondary mathematics teachers perceive statistics as difficult. *Journal of Statistics Education*, *21*(3).

Lee, H., & Hollebrands, K. (2008). Preparing to teach data analysis and probability with technology. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education: Proceedings of the ICMI study 18 and 2009 IASE round table conference.*

Lee, H., & Hollebrands, K. (2011). Characterising and developing teachers' knowledge for teaching statistics with technology. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher Education: A joint ICMI/IASE study.*

Lesser, L., & Groth, R. (2008). *Technological pedagogical content knowledge in statistics.* Paper presented at the 20th International Conference of Technology in Collegiate Mathematics, San Antonio.

Lewis, C., Perry, R., Friedkin, S., & Roth, J. (2012). Improving teaching does improve teachers: Evidence from lesson study. *Journal of Teacher Education*, *63*(5), 368-375.

Lyublinskaya, I., & Tournaki, E. (2012). The effects of teacher content authoring on TPACK and on student achievement in algebra: Research on instruction with the TI-Nspire handheld. In R. Ronau, C. Rakes, & M. Niess (Eds.), *Educational technology, teacher knowledge, and classroom impact: A Research handbook on frameworks and approaches.* (pp. 295-322) Hershey, PA: IGI Global.

Margerum-Leys, J., & Marx, R. W. (2002). Teacher knowledge of educational technology: A case study of student/mentor pairs. *Journal of Educational Computing Research*, *26*, 427 - 462.

Mishra, P., & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017-1054.

Mishra, P., & Koehler, M. (2008). *Introducing technological pedagogical content knowledge*. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, NY.

Morris, A., & Hiebert, J. (2011). Creating shared instructional products: An alternative approach to improving teaching. *Educational Researcher*, *40*(1), 5-14.

Mudzimiri, Rejoice. (2012). *A study of the development of technological pedagogical content knowledge (TPACK) in pre-service secondary mathematics teachers* (Doctoral dissertation). Montana State University, ProQuest, UMI Dissertations Publishing, 2012. 3523442.

National Committee on Mathematical Requirements. (1923). *The Reorganization of Mathematics in Secondary Education*. Washington, D.C.: Mathematical Association of America.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2014). *Principles to Action Report*. Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010a). *Common Core State Standards for Mathematics*. Washington, DC: Authors.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010b). *Appendix A of Common Core State Standards for Mathematics*. Washington, DC: Authors.

Niess, M. L. (2005). Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. *Teaching and Teacher Education*, *21*, 509 - 523.

Niess, M., Browning, C., Driskell, S., Johnston, C., & Harrington, R. (2009). Mathematics Teacher TPACK Standards and Revising Teacher Preparation. In I. Gibson et al. (Eds.), *Proceedings of society for information technology & teacher education international conference 2009*. Chesapeake, VA: Association for the Advancement of Computing in Education.

Niess, M., Sadri, P., & Lee, K. (2007). *Dynamic spreadsheets as learning technology tools: Developing teachers' technology pedagogical content knowledge (TPCK).* Paper presented at the meeting of the American Educational Research Association Annual Conference, Chicago, IL.

Niess, M. L., Suharwoto, G., Lee, K., & Sadri, P. (2006). *Guiding inservice mathematics teachers in developing technology pedagogical content knowledge (TPCK)*. Paper presented at the international conference of Society for Information Technology & Teacher Education (SITE) Conference (March 20th-24th), Orlando, FL.

Peck, R., Gould, R., Miller, S., Wilson, P., & Zbiek, R. (2013). *Developing Essential Understanding of Statistics for Teaching Mathematics in Grades 9-12.* Reston, VA: National Council of Teachers of Mathematics.

Peck, R., Kader, G., & Franklin, C. (2008). Shaping K-12 statistics education in the United States. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE study: Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICMI/IASE study.*

Pierson, M. E. (2001). Technology integration practices as a function of pedagogical expertise. *Journal of Research on Computing in Education*, *33*, 413 - 430.

Pratt, D., Davies, N., & Connor, D. (2011). The role of technology in teaching and learning statistics. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education: A joint ICMI/IASE study.*

Ramsey, F., & Schafer, D. (2013). *The statistical sleuth: A course in methods of data analysis*. Boston: Brooks/Cole, Cengage Learning.

Rogers, E. (1995). *Diffusion of innovations.* New York, Free Press.

Rossman, A., & Franklin, C. (2013). Interview with Christine Franklin. *Journal of Statistics Education*, *21*(3).

Rossman, A., Chance, B., & Medina, E. (2006). Some important considerations between statistics and mathematics and why teachers should care. In G. Burrill & P. Elliot (Eds.), *Thinking and reasoning with data and chance (68th annual yearbook of the National Council of Teachers of Mathematics).* Reston, VA: NCTM.

Rumsey, D. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*(3).

Savin-Baden, M., and Major, C. (2013). *Qualitative research: The essential guide to theory and practice.* London: Routledge.

Scheaffer, R. (2006). Statistics and mathematics: On making a happy marriage. In G. Burrill (Ed.), *Thinking and Reasoning with Data and Chance: 68th NCTM Yearbook (2006)* (p 309-321). Reston: National Council of Teachers of Mathematics.

Scheaffer, R., & Jacobbe, T. (2014). Statistics education in the K-12 schools of the United States: A brief history. *Journal of Statistics Education*, *22*(2).

Schmidt, D., Baran E., Thompson, A., Mishra, P., Koehler, M., & Shin, T. (2009). Technological pedagogical content knowledge (TPACK): The development and validation of an assessment instrument for preservice teachers. *Journal of Research on Technology in Education*, *42*(2), 123-149.

Shulman, L. (1986). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*, 1-22.

Shulman, L. (1987). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*, 4-14.

Sorto, M., & Lesser, L. (2010). Towards measuring technological pedagogical content knowledge in statistics: Middle school teachers using graphing calculators. In *Proceedings of the 2009 IASE Satellite Conference on Next Steps in Statistics Education.*

Stigler, J., & Thompson, B. (2009). Thoughts on creating, accumulating, and utilizing shareable knowledge to improve teaching. *The Elementary School Journal*, *109*(5), 442-457.

Tishkovskaya, S., & Lancaster, G. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, *20*(2), 1-55.

Tournaki, N., & Lyublinskaya, I. (2014). Preparing special education teachers for teaching mathematics and science with technology by integrating the TPACK framework into the curriculum: A study of teachers' perceptions. *Journal of Technology and Teacher Education*, *22*(2), 243-259.

U.S. Department of Education. (2016). *The 2016 national education technology plan: Future reading learning. Reimaging the role of technology in education*. Washington, DC: Office of Educational Technology.

Usiskin, Z. (2015). The relationships between statistics and other subjects in the K-12 curriculum. *Chance, 28*(3).

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

Utts, J. (2010). Unintentional lies in the media: Don't blame journalists for what we don't teach. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eight International Conference on Teaching Statistics* (ICOTS8, July, 2010). *Ljubljana, Slovenia.* Voorburg, The Netherlands: International Statistical Institute.

Voogt, J., Fisser, P., Roblin N., Tondeur, J., & van Braak, J. (2012). Technological pedagogical content knowledge - a review of the literature. *Journal of Computer Assisted Learning*, *29*, 109-121.

Wachira, P., & Keengwe, J. (2011). Technology integration barriers: Urban school mathematics teachers' perspectives. *Journal of Science Education & Technology*, *20*(1), 17-25.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

APPENDICES

APPENDIX A

ANNOTATED STATISTICS UNIT

Statistics Unit

Unit Overview:

The overall theme of this unit revolves around music playlists and listening to music while studying. Lesson 1 is an introduction to randomness, an important concept in statistics. Students will explore whether a given dataset displays the characteristics of randomness. Lesson 2 focuses on the normal distribution and when it is an appropriate model to use. Students will be asked to use the normal distribution by solving unknown percentage problems and unknown value problems. Lesson 3 provides students the opportunity to explore the differences between observational studies and experiments and understand the advantages of randomized experiments. Lesson 4 centers on interval estimation and margin of error, allowing students to evaluate the quality of data by examining the width of confidence intervals. Finally, Lesson 5 uses data from an experiment to compare two different treatments.

About the Lesson Plans:

Each lesson plan contains two things: (1) an overview and (2) detailed learning plan. Each lesson centers on particular statistics content standards and learning goals, and has the potential to guide and inform all facets of the lesson.

The overview of each lesson is designed to provide basic information, such as topics to be taught, prior knowledge needed, activities used, means of assessment, estimated time of completion, learning objectives, teaching materials, and the primary statistics content standards and mathematical practices emphasized. The content standards were purposely chosen by selecting those recommended to be taught in Algebra II/Mathematics III as stated in *Appendix A: Designing High School Mathematics Courses Based on the Common Core State Standards* (pages 43 and 74).

The learning plans are very detailed and contain two main pieces of information: the teacher activity and estimated time. Each learning plan is divided into segments as a means to organize how information is presented. The detail of the learning plans is intentional and meant to provide enough information to implement the concepts in the classroom. For example, suggested plots, questions (and answers), and terminology are provided under the assumption that there are no additional resources to reference. It is intended that the level of detail will allow all teachers to implement the concepts in such a way that they fully address the content standards.

About the Annotations in the Lesson Plans:

There are special annotations written within Lesson 3 (*Observational Studies and Experiments*) and Lesson 4 (*Interval Estimation and Margin of Error*). These annotations are denoted by green font and are surrounded by a border and were created and written with the guidance of various statistics experts. The purpose of the annotations is to ease implementation of the lesson. These annotations provide generally useful information such as summaries of key concepts, reasons for instructional decisions, anticipated student responses, misconceptions, and difficulties often encountered during the lessons, and additional questions to ask and plots to create that can help deepen students' conceptual understanding of the material. In addition, the annotations provide suggestions and hints regarding the use of technology during the lesson. For instance, key components of the web applet (that might not be readily known such as clicking on a data point to reveal more information about it, resetting the plots in the applet, etc.) are described. Furthermore, instructions on how to use the web applet and what your screen should look like are provided within the annotations.

Technology-Based Activities:

Lessons 3 and 4 contain activities (and solutions) that use a web applet which can be used on computers, tablets, and smartphones. The main link to the web applet is: https://msustat.shinyapps.io/SpIntro-Stats/ and the backup link is: http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/.

Sample Homework Problems:
For each lesson, sample homework problems and solutions are provided. These problems are designed to give students additional opportunities to practice the concepts they learned throughout the lesson.

Sample Summative Assessments:
Sample test questions (with solutions) from Lessons 2 – 5 are also provided for your convenience. In addition, an optional poster project description is provided. This project provides students an opportunity to apply what they have learned throughout the entire unit. Students will formulate a question, collect and analyze appropriate data, draw a conclusion, and communicate their process and results in a poster.

CCSSM Standards Addressed:

- *S.ID.4.* Use the mean and standard deviation of a data set to fit a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.
- *S.IC.1.* Understand statistics as a process for making inferences about population parameters based on a random sample from that population.
- *S.IC.2.* Decide if a specified model is consistent with results from a given data-generating process.
- *S.IC.3.* Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.
- *S.IC.4.* Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.
- *S.IC.5.* Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

The following three standards may or may not be addressed in the lessons. For instance, in Lesson 1, students typically make decisions about the playlists using probability concepts and therefore, S.MD.7 would be incorporated into that lesson.

- *S.IC.6.* Evaluate reports based on data.
- *S.MD.6.* Use probabilities to make fair decisions.
- *S.MD.7.* Analyze decisions and strategies using probability concepts.

CCSSM Mathematical Practices Emphasized:

- *MP1.* Make sense of problems and persevere in solving them.
- *MP3.* Construct viable arguments and critique the reasoning of others.
- *MP4.* Model with mathematics.
- *MP5.* Use appropriate tools strategically.

Summary of Lessons:

Lesson/Topic	Summary
Lesson 1 (Randomness)	This lesson centers on the iPod Shuffle, a model-eliciting activity that teaches students to recognize characteristics of randomly sampled or randomly generated data. Students are given three samples of data: a set of 25 randomly generated playlists for students to use as a basis for describing characteristics of a randomly generated playlist, a set of five additional randomly-generated playlists on which to test their rules, and the set of Mr. Hoffman's original three playlists (which students are asked to judge based on their rules). Each group will write a brief memo describing their decisions. This activity serves as an introduction to the idea of randomness, an important concept in statistics. *Note: This is not an activity where we talk about random assignment or random selection, but rather how difficult it is for us, as humans, to identify sequences that are truly random.*
Lesson 2 (Normal Distributions)	This lesson centers on the normal distribution, why it is important, and when it is appropriate to use it. A brief review of the normal distribution including the basic properties of the normal distribution will be presented first. Students will learn that if a distribution is normal, then the mean and standard deviation can tell us everything else about the distribution. The normal distribution allows us to find probabilities (areas under a curve) and students will solve two related problems: the unknown percentage problem and the unknown value problem. Furthermore, students will explore two datasets (song lengths and album prices) and recognize under which dataset the normal model is more appropriate.
Lesson 3 (Observational Studies and Experiments)	This lesson centers on observational studies, experiments, and the advantages of a randomized experiment. A very brief review of the difference between categorical and quantitative variables should be presented if needed. The remainder of the lesson is an activity that first guides students through the definitions of an observational study and an experiment and then uses a web applet to help students learn the advantages of a randomized experiment. The main experiment proposed focuses on investigating the effectiveness of studying with music using a treatment and control group. The concepts of lurking variables and how randomization (i.e., randomly assigning treatments) "evens out" the lurking variable are also explored throughout this activity. *Note: This web applet works on computers, tablets, and smartphones.*
Lesson 4 (Interval Estimation and Margin of Error)	This lesson centers on interval estimation and margin of error. The activity uses data from a sample survey and students are asked to use information to estimate a population proportion. Next, the activity guides students through the idea of variability within sample proportions and how an interval estimate (opposed to a point estimate) can describe more of a "story." A web applet is used to conduct a simulation to investigate how variable a sample proportion might be. Students are introduced to the concept of bootstrapping and run a simulation to develop a margin

	of error and interval estimates. Finally, 95% confidence intervals are constructed using the web app ("percentile method") and by hand ("Estimate $\pm$ 2 $\times$ SE method"). *Note: This web applet works on computers, tablets, and smartphones.*
Lesson 5 (Comparing Two Treatments)	This lesson centers on comparing two treatments from a randomized experiment. In this experiment, 60 hypothetical randomly selected students are randomly assigned to either a treatment group (music playing) or control group (silence). The main research question is "Does listening to music improve SAT scores?" Students are guided through the process of conducting a test to make inferences.

Overview of Lesson 1

<table>
<tr><td colspan="2">Topic of Lesson:
Randomness</td><td>Date:</td></tr>
<tr><td colspan="3">Summary of Lesson:
This lesson centers on the iPod Shuffle, a model-eliciting activity that teaches students to recognize characteristics of randomly sampled or randomly generated data. Students are given three samples of data: a set of 25 randomly generated playlists for students to use as a basis for describing characteristics of a randomly generated playlist, a set of five additional randomly-generated playlists on which to test their rules, and the set of Mr. Hoffman's original three playlists (which students are asked to judge based on their rules). Each group will write a brief memo describing their decisions. This activity serves as an introduction to the idea of randomness, an important concept in statistics.

Note: This is not an activity where we talk about random assignment or random selection, but rather how difficult it is for us, as humans, to identify sequences that are truly random.</td></tr>
<tr><td>Prerequisites:
None</td><td>Activities:
How Random is the iPod’s Shuffle?</td><td>Estimated Time:
Approximately 50 minutes</td></tr>
<tr><td colspan="2">Learning Objectives
Students will be able to …
• identify characteristics of randomly generated data.
• examine data from a real-world problem.</td><td>Assessments
• Optional HW assignment with problems of your choice – (see randomness_hw_sample.docx for possible homework problems)</td></tr>
<tr><td colspan="2">CCSSM Standards Addressed
• S.IC.2. Decide if a specified model is consistent with results from a given data-generating process.</td><td>CCSSM Mathematical Practices Emphasized
• MP1. Make sense of problems and persevere in solving them.
• MP3. Construct viable arguments and critique the reasoning of others.</td></tr>
<tr><td colspan="3">Teaching Materials
• Copies of the How Random is the iPod’s Shuffle activity (one per group or one per student)
• Copies of the following playlists (one per group or one per student):
o 25_random_playlists.pdf
o 5_random_playlists.pdf
o Hoffman_playlists.pdf</td></tr>
</table>

Learning Plan for Lesson 1

Teacher Activity	Estimated Time
1. *Preparation/Introduction* Give students the *How Random is the iPod's Shuffle* activity. As an introduction to this activity, students should read the short introduction to the activity and discuss/answer problems #1 – 3.	~ 5-10 min.
2. *Group Task* Each group member should read the group task description and discuss the task (if necessary) with their group members. **Part 1: Explore and Describe** Give each group one set of 25 randomly generated playlists (*25_random_playlists.pdf* file). Each group should write at least three characteristics of a randomly generated playlist. *Note: Each playlist is truly random and independent of one another.* **Part 2: Develop Rules** Students work together to develop rules that flag playlists that do **NOT** appear to be randomly generated. Each characteristic they mention in Part 1 should be incorporated into a rule. These rules should be clearly defined so that other individuals could follow and use their rules. *Note: Students should be looking at **ARTISTS**, not songs because Mr. Hoffman's main concern was that some **artists** are being played too often. You may have to remind students of this point.* **Part 3: Test and Refine Rules** Once students have developed their rules, they should let you know they are ready to test them. Give each group one additional set of 5 randomly generated playlists (*5_random_playlists.pdf* file) created by Tim Cook (Apple CEO). Each group should use their rules on this additional set of playlists and refine them if necessary. *Note: Each of these playlists is truly random and independent of one another. Ideally, students should not see any violations in their rules as these playlists are randomly generated.* **Part 4: Evaluate**	~ 30-35 min.

After students have tested and refined (if necessary) their rules, give each group a set of Mr. Hoffman's three playlists (*Hoffman_playlists.pdf* file). Students are to use their rules to find if there is "convincing" evidence that Mr. Hoffman's shuffle feature is producing playlists that are **NOT** randomly generated. **Part 5: Summarize** Each group is to write a short memo to Mr. Hoffman explaining their decision about his iPod's shuffle feature. They should explain their rules and what they found in regards to his playlists.	
3. *Wrap-Up* **Discussion:** Discuss the different rules and solutions students came up with. Some questions to ask students include: ➢ What did you learn about random sequences? ➢ What do you now know or believe that you did not know at the start of this activity? ➢ What was your process for coming up with rules? ➢ How might your rules have changed if there weren't an equal number of songs for each artist? Or a longer set of songs per playlist?	~ 5 min

Take-Home Messages ➢ As humans, the ability to detect randomness is difficult. We often convince ourselves that we see patterns in sequences even when there are none.	
4. *Homework* (OPTIONAL) At your discretion, assign homework problems. The questions presented in the *randomness_hw_sample.docx* file give students the opportunity to read more about randomness (via a blog) and further develop their ideas about it by answering a few questions. Solutions to these sample problems can be found in the *randomness_hw_sample_solutions.docx* file	



How Random is the iPod's Shuffle?



A "shuffle" feature built into the iPod software takes a list of songs, called a playlist, and rearranges them in a random order. How would you know if your sequence of songs is random? Randomness is a key concept in statistics, but how good are we at detecting randomness? We will investigate this idea today.

With your group briefly discuss and write your responses to the following questions.

1. Do you have an iPod or some other digital music player? Have you used the shuffle feature? If you have used the shuffle feature, have you ever wondered how truly random it is?

2. What comes to mind when you hear the word, "random"?

3. If the iPod shuffle feature is ***not*** producing a random sequence of songs, then what might the sequence of songs look like? What would you expect to see?

Group Task

Albert Hoffman, an iPod owner, has written a letter to Apple to complain about the iPod Shuffle feature. He writes that everyday he takes an hour-long walk and listens to his iPod using the shuffle feature. **He believes that the shuffle feature is producing playlists in which *some artists are played too often and others are not played enough*.**

He has claimed that the iPod Shuffle feature is ***not*** generating random playlists. As evidence, Mr. Hoffman has provided both his music library (8 artists with 10 songs each) and three playlists (20 songs each) that his iPod generated using the shuffle feature.

Tim Cook, the CEO of Apple, Inc., has contacted your group to respond to Mr. Hoffman's complaint. He has provided your group with several playlists of 20 songs each using the same songs as Mr. Hoffman's library but generating them using a genuine random number generation method.

To help your group respond to Mr. Hoffman, the next four sections are designed to help your group explore properties of the randomly generated lists to develop rules that could help determine whether a set of playlists provide evidence that the shuffle feature is ***not*** producing randomly selected songs. The fifth section asks you to write a memo to Mr. Hoffman addressing his concerns.

1. Explore and Describe

Examine the randomly generated playlists (your group will be given 25) to get an idea of the characteristics of these random lists. Write down three characteristics of a randomly generated playlist.

2. Develop Rules

Use the set of characteristics that your group wrote down to *create a set of rules* that help identify playlists that **do not appear to have been randomly generated**. (Be sure that each of your characteristics is included in a rule.) *These rules should be clearly stated so that another person could easily use them.*

3. Test and Refine Rules

Your group will now be given five additional playlists created by Tim Cook that are truly randomly-generated playlists on which to test your rules. Let you teacher know that you are ready to receive these playlists. See whether the set of rules your group generated would lead someone to (incorrectly) question whether these playlists are not randomly generated. Then, based on the performance of your group's set of rules, adapt or change them, as your group feels necessary and write them down.

4. Evaluate

Your group will now be provided with Mr. Hoffman's original three playlists. Let your teacher know that you are ready to receive these playlists. Apply your group's rules to these three playlists to judge whether there is convincing evidence that Mr. Hoffman's iPod Shuffle feature is producing playlists in which some ***artists*** are played too often and others are not played enough (i.e., the playlists do not seem to be randomly generated).

5. Summarize

Respond to Mr. Hoffman's claim that the shuffle feature is not random by writing a short paragraph about your group's findings.



How Random is the iPod's Shuffle?
Solutions



With your group briefly discuss and write your responses to the following questions.

1. Do you have an iPod or some other digital music player? Have you used the shuffle feature? If you have used the shuffle feature, have you ever wondered how truly random it is?

   Answers will vary.

2. What comes to mind when you hear the word, "random"?

   Answers will vary.

   Some possibilities include:
   - *no pattern*
   - *chaotic*
   - *no structure*
   - *by chance*
   - *no order*

3. If the iPod shuffle feature is ***not*** producing a random sequence of songs, then what might the sequence of songs look like? What would you expect to see?

   Answers will vary.

   Some possibilities include:
   - *same song played multiple times in a sequence.*
   - *same artist played over and over again even though there are many artists on the playlists.*

Group Task

1. **Explore and Describe**

   Examine the randomly generated playlists (your group will be given 25) to get an idea of the characteristics of these random lists. Write down three characteristics of a randomly generated playlist.

   Answers will vary.

   Some characteristics may involve:
   - *number of songs on a playlist by the same artist (e.g., on a randomly generated playlist, each artist only appeared 2-3 times (i.e., 20 songs on a playlist and 8 artists in the music library → each artist appears 2-3 times (20/8=2.5))*
   - *number of songs in a row by the same artists*
   - *number of different artists on a playlist*

2. **Develop Rules**

   Use the set of characteristics that your group wrote down to *create a set of rules* that help identify playlists that **do not appear to have been randomly generated**. (Be sure that each of your characteristics is included in a rule.) *These rules should be clearly stated so that another person could easily use them.*

   Answers will vary.

   Some of the rules may involve
   - *range of numbers of songs on a playlist by the same artist (e.g., NOT random if an artist has 5 songs on a playlist)*
   - *number of songs in a row by the same artists*
   - *number of different artists on a playlist*

3. **Test and Refine Rules**

   Your group will now be given five additional playlists created by Tim Cook that are truly randomly generated playlists on which to test your rules. Let you teacher know that you are ready to receive these playlists. See whether the set of rules your group generated would lead someone to (incorrectly) question whether these playlists are not randomly generated. Then, based on the performance of your group's set of rules, adapt or change them, as your group feels necessary and write them down.

   Answers will vary.

In my experience, there have been slight changes to the rules and most of these changes have involved altering the numbers students have set in their rules (e.g., NOT random if an artist has more than 3 songs on the playlist).

4. Evaluate

Your group will now be provided with Mr. Hoffman's original three playlists. Let your teacher know that you are ready to receive these playlists. Apply your group's rules to these three playlists to judge whether there is convincing evidence that Mr. Hoffman's iPod Shuffle feature is producing playlists in which some ***artists*** are played too often and others are not played enough (i.e., the playlists do not seem to be randomly generated).

Students may write down some notes/findings as they analyze Mr. Hoffman's playlists. For example, one group used this space to keep track of artists and tallied the number of times an artist appeared in each playlist. Other groups may write on the playlists themselves.

5. Summarize

Respond to Mr. Hoffman's claim that the shuffle feature is not random by writing a short paragraph about your group's findings.

Answers will vary.

The memos ought to include:

- *characteristics of randomly generated playlists.*
- *clear and detailed rules that they developed.*
- *A clear and direct answer to Mr. Hoffman explaining their findings. Did they find that his iPod shuffle feature was working correctly or incorrectly?*
    - *A detailed explanation of their findings. For example, did they find any violations of the rules when analyzing Mr. Hoffman's playlists? If so, what were they?*

5 Random Playlists

(Set 2) Track	Artist	Album Year
Mayfly	Belle & Sebastian	If You're Feeling Sinister 1996
My Baby	Janis Joplin	Pearl 1971
Rehearsals For Departure	Damien Jurado	Rehearsals For Departure 1999
One More Dollar	Gillian Welch	Revival 1996
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
Immigrant Song	Led Zeppelin	Led Zeppelin III 1970
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996
Only One And Only	Gillian Welch	Revival 1996
I'll Be Around	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
Gallows Pole	Led Zeppelin	Led Zeppelin III 1970
Shoot To Thrill	AC/DC	Back In Black 1980
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
Pass You By	Gillian Welch	Revival 1996
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Rock And Roll Ain't Noise Poll	AC/DC	Back In Black 1980
Hells Bells	AC/DC	Back In Black 1980
Trust Me	Janis Joplin	Pearl 1971
Curbside	Damien Jurado	Rehearsals For Departure 1999
Acony Bell	Gillian Welch	Revival 1996

(Set 2) Track	Artist	Album Year
Trust Me	Janis Joplin	Pearl 1971
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Hells Bells	AC/DC	Back In Black 1980
Half Moon	Janis Joplin	Pearl 1971
Curbside	Damien Jurado	Rehearsals For Departure 1999
Buried Alive In The Blues	Janis Joplin	Pearl 1971
Only One And Only	Gillian Welch	Revival 1996
Get Me Away From Here, I'm Dying	Belle & Sebastian	If You're Feeling Sinister 1996
Mercedes Benz	Janis Joplin	Pearl 1971
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister 1996
Treefingers	Radiohead	Kid A 2000
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
The Ending	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Eyes For Windows	Damien Jurado	Rehearsals For Departure 1999
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
Love The Same	Damien Jurado	Rehearsals For Departure 1999
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
Tornado	Damien Jurado	Rehearsals For Departure 1999
Dig Down	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Rock And Roll Ain't Noise Poll	AC/DC	Back In Black 1980

(Set 2) Track	Artist	Album Year
Buried Alive In The Blues	Janis Joplin	Pearl 1971
Everything In It's Right Place	Radiohead	Kid A 2000
I'll Be Around	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Treefingers	Radiohead	Kid A 2000
Mercedes Benz	Janis Joplin	Pearl 1971
The Ending	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Half Moon	Janis Joplin	Pearl 1971
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
A Woman Left Lonely	Janis Joplin	Pearl 1971
Hats Off To (Roy) Harper	Led Zeppelin	Led Zeppelin III 1970
Optimistic	Radiohead	Kid A 2000
Ohio	Damien Jurado	Rehearsals For Departure 1999
Love The Same	Damien Jurado	Rehearsals For Departure 1999
What Do You Do For Money Honey	AC/DC	Back In Black 1980
Get It While You Can	Janis Joplin	Pearl 1971
Hells Bells	AC/DC	Back In Black 1980
Cry Baby	Janis Joplin	Pearl 1971
Honey Baby	Damien Jurado	Rehearsals For Departure 1999
Idioteque	Radiohead	Kid A 2000

(Set 2) Track	Artist	Album Year
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
By The Mark	Gillian Welch	Revival 1996
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996
Barroom Girls	Gillian Welch	Revival 1996
How To Disappear Completely	Radiohead	Kid A 2000
Get It While You Can	Janis Joplin	Pearl 1971
Orphan Girl	Gillian Welch	Revival 1996
Gallows Pole	Led Zeppelin	Led Zeppelin III 1970
Kid A	Radiohead	Kid A 2000
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister 1996
Idioteque	Radiohead	Kid A 2000
Get Me Away From Here, I'm Dying	Belle & Sebastian	If You're Feeling Sinister 1996
Rehearsals For Departure	Damien Jurado	Rehearsals For Departure 1999
Morning Bell	Radiohead	Kid A 2000
Eyes For Windows	Damien Jurado	Rehearsals For Departure 1999
You Shook Me All Night Long	AC/DC	Back In Black 1980
Celebration Day	Led Zeppelin	Led Zeppelin III 1970
What Do You Do For Money Honey	AC/DC	Back In Black 1980
The National Anthem	Radiohead	Kid A 2000
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002

(Set 2) Track	Artist	Album Year
Acony Bell	Gillian Welch	Revival 1996
Mehan	Bobby Bare Jr.	Young Criminal's Starvation League 2002
How To Disappear Completely	Radiohead	Kid A 2000
Get Me Away From Here, I'm Dying	Belle & Sebastian	If You're Feeling Sinister 1996
Saturday	Damien Jurado	Rehearsals For Departure 1999
Morning Bell	Radiohead	Kid A 2000
Gallows Pole	Led Zeppelin	Led Zeppelin III 1970
The Monk At The Disco	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Tangerine	Led Zeppelin	Led Zeppelin III 1970
Idioteque	Radiohead	Kid A 2000
Everything In It's Right Place	Radiohead	Kid A 2000
Kid A	Radiohead	Kid A 2000
If You're Feeling Sinister	Belle & Sebastian	If You're Feeling Sinister 1996
Immigrant Song	Led Zeppelin	Led Zeppelin III 1970
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
What Do You Do For Money Honey	AC/DC	Back In Black 1980
Curbside	Damien Jurado	Rehearsals For Departure 1999
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996
Move Over	Janis Joplin	Pearl 1971

Hoffman's Playlists

Track	Artist	Album Year
Out On The Tiles	Led Zeppelin	Led Zeppelin III 1970
Tornado	Damien Jurado	Rehearsals For Departure 1999
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
If You're Feeling Sinister	Belle & Sebastian	If You're Feeling Sinister 1996
Acony Bell	Gillian Welch	Revival 1996
Only One And Only	Gillian Welch	Revival 1996
One More Dollar	Gillian Welch	Revival 1996
Celebration Day	Led Zeppelin	Led Zeppelin III 1970
That's The Way	Led Zeppelin	Led Zeppelin III 1970
In Limbo	Radiohead	Kid A 2000
Everything In It's Right Place	Radiohead	Kid A 2000
Tragedy	Damien Jurado	Rehearsals For Departure 1999
Since I've Been Loving You	Led Zeppelin	Led Zeppelin III 1970
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Ohio	Damien Jurado	Rehearsals For Departure 1999
Immigrant Song	Led Zeppelin	Led Zeppelin III 1970
Pass You By	Gillian Welch	Revival 1996
Back In Black	AC/DC	Back In Black 1980
Shake A Leg	AC/DC	Back In Black 1980
Saturday	Damien Jurado	Rehearsals For Departure 1999

Track	Artist	Album Year
A Woman Left Lonely	Janis Joplin	Pearl 1971
Shake A Leg	AC/DC	Back In Black 1980
Me & Bobby McGee	Janis Joplin	Pearl 1971
The National Anthem	Radiohead	Kid A 2000
Hats Off To (Roy) Harper	Led Zeppelin	Led Zeppelin III 1970
Celebration Day	Led Zeppelin	Led Zeppelin III 1970
How To Disappear Completely	Radiohead	Kid A 2000
Half Moon	Janis Joplin	Pearl 1971
Curbside	Damien Jurado	Rehearsals For Departure 1999
Treefingers	Radiohead	Kid A 2000
Rehearsals For Departure	Damien Jurado	Rehearsals For Departure 1999
Tangerine	Led Zeppelin	Led Zeppelin III 1970
Kid A	Radiohead	Kid A 2000
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Hells Bells	AC/DC	Back In Black 1980
Trust Me	Janis Joplin	Pearl 1971
I'll Be Around	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Flat Chested Girl From Maynardville	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
My Baby	Janis Joplin	Pearl 1971

Track	Artist	Album Year
Get It While You Can	Janis Joplin	Pearl 1971
Pass You By	Gillian Welch	Revival 1996
Back In Black	AC/DC	Back In Black 1980
I'll Be Around [Reprise]	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The National Anthem	Radiohead	Kid A 2000
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister 1996
Love The Same	Damien Jurado	Rehearsals For Departure 1999
In Limbo	Radiohead	Kid A 2000
You Shook Me All Night Long	AC/DC	Back In Black 1980
Dig Down	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Gallows Pole	Led Zeppelin	Led Zeppelin III 1970
Out On The Tiles	Led Zeppelin	Led Zeppelin III 1970
Since I've Been Loving You	Led Zeppelin	Led Zeppelin III 1970
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
Celebration Day	Led Zeppelin	Led Zeppelin III 1970
Idioteque	Radiohead	Kid A 2000
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
Optimistic	Radiohead	Kid A 2000

25 Random Playlists

(Set 1) Track	Artist	Album Year
The National Anthem	Radiohead	Kid A 2000
Hells Bells	AC/DC	Back In Black 1980
Immigrant Song	Led Zeppelin	Led Zeppelin III 1970
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996
Back In Black	AC/DC	Back In Black 1980
Eyes For Windows	Damien Jurado	Rehearsals For Departure 1999
Trust Me	Janis Joplin	Pearl 1971
Gallows Pole	Led Zeppelin	Led Zeppelin III 1970
Ohio	Damien Jurado	Rehearsals For Departure 1999
Flat Chested Girl From Maynardville	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister 1996
Buried Alive In The Blues	Janis Joplin	Pearl 1971
Me & Bobby McGee	Janis Joplin	Pearl 1971
Tornado	Damien Jurado	Rehearsals For Departure 1999
Mayfly	Belle & Sebastian	If You're Feeling Sinister 1996
How To Disappear Completely	Radiohead	Kid A 2000
That's The Way	Led Zeppelin	Led Zeppelin III 1970
My Baby	Janis Joplin	Pearl 1971
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Idioteque	Radiohead	Kid A 2000

(Set 1) Track	Artist	Album Year
Idioteque	Radiohead	Kid A 2000
Cry Baby	Janis Joplin	Pearl 1971
Tear My Stillhouse Down	Gillian Welch	Revival 1996
Ohio	Damien Jurado	Rehearsals For Departure 1999
Saturday	Damien Jurado	Rehearsals For Departure 1999
Only One And Only	Gillian Welch	Revival 1996
Treefingers	Radiohead	Kid A 2000
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996
Shoot To Thrill	AC/DC	Back In Black 1980
Annabelle	Gillian Welch	Revival 1996
Motion Picture Soundtrack [Acoustic]	Radiohead	Kid A 2000
I'll Be Around [Reprise]	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister 1996
Flat Chested Girl From Maynardville	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
Me & Bobby McGee	Janis Joplin	Pearl 1971
Back In Black	AC/DC	Back In Black 1980
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Dig Down	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Mayfly	Belle & Sebastian	If You're Feeling Sinister 1996

(Set 1) Track	Artist	Album Year
Immigrant Song	Led Zeppelin	Led Zeppelin III 1970
Given The Dog A Bone	AC/DC	Back In Black 1980
In Limbo	Radiohead	Kid A 2000
Out On The Tiles	Led Zeppelin	Led Zeppelin III 1970
You Shook Me All Night Long	AC/DC	Back In Black 1980
Pass You By	Gillian Welch	Revival 1996
Me & Bobby McGee	Janis Joplin	Pearl 1971
Cry Baby	Janis Joplin	Pearl 1971
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
One More Dollar	Gillian Welch	Revival 1996
Gallows Pole	Led Zeppelin	Led Zeppelin III 1970
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Mehan	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Mercedes Benz	Janis Joplin	Pearl 1971
Let Me Put My Love Into You	AC/DC	Back In Black 1980
Buried Alive In The Blues	Janis Joplin	Pearl 1971

(Set 1) Track	Artist	Album Year
A Woman Left Lonely	Janis Joplin	Pearl 1971
Morning Bell	Radiohead	Kid A 2000
Eyes For Windows	Damien Jurado	Rehearsals For Departure 1999
Orphan Girl	Gillian Welch	Revival 1996
The Ending	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Back In Black	AC/DC	Back In Black 1980
My Baby	Janis Joplin	Pearl 1971
Half Moon	Janis Joplin	Pearl 1971
Rehearsals For Departure	Damien Jurado	Rehearsals For Departure 1999
Buried Alive In The Blues	Janis Joplin	Pearl 1971
Celebration Day	Led Zeppelin	Led Zeppelin III 1970
What Do You Do For Money Honey	AC/DC	Back In Black 1980
Treefingers	Radiohead	Kid A 2000
How To Disappear Completely	Radiohead	Kid A 2000
Optimistic	Radiohead	Kid A 2000
Immigrant Song	Led Zeppelin	Led Zeppelin III 1970
Kid A	Radiohead	Kid A 2000
Annabelle	Gillian Welch	Revival 1996
Tear My Stillhouse Down	Gillian Welch	Revival 1996
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996

(Set 1) Track	Artist	Album Year
In Limbo	Radiohead	Kid A 2000
Annabelle	Gillian Welch	Revival 1996
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
Have A Drink On Me	AC/DC	Back In Black 1980
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996
Treefingers	Radiohead	Kid A 2000
Only One And Only	Gillian Welch	Revival 1996
Dig Down	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Mercedes Benz	Janis Joplin	Pearl 1971
Ohio	Damien Jurado	Rehearsals For Departure 1999
One More Dollar	Gillian Welch	Revival 1996
Barroom Girls	Gillian Welch	Revival 1996
I'll Be Around [Reprise]	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Cry Baby	Janis Joplin	Pearl 1971
Get Me Away From Here, I'm Dying	Belle & Sebastian	If You're Feeling Sinister 1996
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Tragedy	Damien Jurado	Rehearsals For Departure 1999

(Set 1) Track	Artist	Album Year
Hells Bells	AC/DC	Back In Black 1980
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Curbside	Damien Jurado	Rehearsals For Departure 1999
Buried Alive In The Blues	Janis Joplin	Pearl 1971
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
That's The Way	Led Zeppelin	Led Zeppelin III 1970
Tornado	Damien Jurado	Rehearsals For Departure 1999
I'll Be Around	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Trust Me	Janis Joplin	Pearl 1971
Rehearsals For Departure	Damien Jurado	Rehearsals For Departure 1999
Shake A Leg	AC/DC	Back In Black 1980
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
Back In Black	AC/DC	Back In Black 1980
Cry Baby	Janis Joplin	Pearl 1971
Mercedes Benz	Janis Joplin	Pearl 1971
Flat Chested Girl From Maynardville	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Shoot To Thrill	AC/DC	Back In Black 1980
Gallows Pole	Led Zeppelin	Led Zeppelin III 1970
Let Me Put My Love Into You	AC/DC	Back In Black 1980
Motion Picture Soundtrack [Acoustic]	Radiohead	Kid A 2000

(Set 1) Track	Artist	Album Year
Dig Down	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Rock And Roll Ain't Noise Pollution	AC/DC	Back In Black 1980
Friends	Led Zeppelin	Led Zeppelin III 1970
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
Get Me Away From Here, I'm Dying	Belle & Sebastian	If You're Feeling Sinister 1996
Tragedy	Damien Jurado	Rehearsals For Departure 1999
Only One And Only	Gillian Welch	Revival 1996
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
The National Anthem	Radiohead	Kid A 2000
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
The Monk At The Disco	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Everything In It's Right Place	Radiohead	Kid A 2000
Tornado	Damien Jurado	Rehearsals For Departure 1999
Love The Same	Damien Jurado	Rehearsals For Departure 1999
Saturday	Damien Jurado	Rehearsals For Departure 1999
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
Kid A	Radiohead	Kid A 2000
A Woman Left Lonely	Janis Joplin	Pearl 1971

(Set 1) Track	Artist	Album Year
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Bron-Yr-Aur Stomp	Led Zeppelin	Led Zeppelin III 1970
Cry Baby	Janis Joplin	Pearl 1971
Like Dylan In The Movies	Belle & Sebastian	If You're Feeling Sinister 1996
Idioteque	Radiohead	Kid A 2000
The National Anthem	Radiohead	Kid A 2000
Given The Dog A Bone	AC/DC	Back In Black 1980
Kid A	Radiohead	Kid A 2000
Tear My Stillhouse Down	Gillian Welch	Revival 1996
Rehearsals For Departure	Damien Jurado	Rehearsals For Departure 1999
Since I've Been Loving You	Led Zeppelin	Led Zeppelin III 1970
My Baby	Janis Joplin	Pearl 1971
Back In Black	AC/DC	Back In Black 1980
Annabelle	Gillian Welch	Revival 1996
Paper Wings	Gillian Welch	Revival 1996
One More Dollar	Gillian Welch	Revival 1996
Eyes For Windows	Damien Jurado	Rehearsals For Departure 1999
Trust Me	Janis Joplin	Pearl 1971

(Set 1) Track	Artist	Album Year
Kid A	Radiohead	Kid A 2000
What Do You Do For Money Honey	AC/DC	Back In Black 1980
The National Anthem	Radiohead	Kid A 2000
Trust Me	Janis Joplin	Pearl 1971
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister 1996
Immigrant Song	Led Zeppelin	Led Zeppelin III 1970
Me & Bobby McGee	Janis Joplin	Pearl 1971
Flat Chested Girl From Maynardville	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Mercedes Benz	Janis Joplin	Pearl 1971
Back In Black	AC/DC	Back In Black 1980
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996
Annabelle	Gillian Welch	Revival 1996
I'll Be Around [Reprise]	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Curbside	Damien Jurado	Rehearsals For Departure 1999
You Shook Me All Night Long	AC/DC	Back In Black 1980
Shake A Leg	AC/DC	Back In Black 1980
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Optimistic	Radiohead	Kid A 2000
Honey Baby	Damien Jurado	Rehearsals For Departure 1999
Cry Baby	Janis Joplin	Pearl 1971

(Set 1) Track	Artist	Album Year
Buried Alive In The Blues	Janis Joplin	Pearl 1971
Mercedes Benz	Janis Joplin	Pearl 1971
Me & Bobby McGee	Janis Joplin	Pearl 1971
By The Mark	Gillian Welch	Revival 1996
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Morning Bell	Radiohead	Kid A 2000
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
A Woman Left Lonely	Janis Joplin	Pearl 1971
Move Over	Janis Joplin	Pearl 1971
Kid A	Radiohead	Kid A 2000
What Do You Do For Money Honey	AC/DC	Back In Black 1980
In Limbo	Radiohead	Kid A 2000
Tragedy	Damien Jurado	Rehearsals For Departure 1999
Tear My Stillhouse Down	Gillian Welch	Revival 1996
Flat Chested Girl From Maynardville	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Motion Picture Soundtrack [Acoustic]	Radiohead	Kid A 2000
Eyes For Windows	Damien Jurado	Rehearsals For Departure 1999
I'll Be Around	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister 1996
Orphan Girl	Gillian Welch	Revival 1996

(Set 1) Track	Artist	Album Year
Cry Baby	Janis Joplin	Pearl 1971
Since I've Been Loving You	Led Zeppelin	Led Zeppelin III 1970
That's The Way	Led Zeppelin	Led Zeppelin III 1970
Celebration Day	Led Zeppelin	Led Zeppelin III 1970
A Woman Left Lonely	Janis Joplin	Pearl 1971
Pass You By	Gillian Welch	Revival 1996
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Have A Drink On Me	AC/DC	Back In Black 1980
Honey Baby	Damien Jurado	Rehearsals For Departure 1999
You Shook Me All Night Long	AC/DC	Back In Black 1980
Rehearsals For Departure	Damien Jurado	Rehearsals For Departure 1999
Idioteque	Radiohead	Kid A 2000
One More Dollar	Gillian Welch	Revival 1996
Given The Dog A Bone	AC/DC	Back In Black 1980
If You're Feeling Sinister	Belle & Sebastian	If You're Feeling Sinister 1996
Mehan	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Optimistic	Radiohead	Kid A 2000
In Limbo	Radiohead	Kid A 2000
Hells Bells	AC/DC	Back In Black 1980
Me & Bobby McGee	Janis Joplin	Pearl 1971

(Set 1) Track	Artist	Album Year
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
Eyes For Windows	Damien Jurado	Rehearsals For Departure 1999
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Treefingers	Radiohead	Kid A 2000
Everything In It's Right Place	Radiohead	Kid A 2000
Mehan	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Optimistic	Radiohead	Kid A 2000
Only One And Only	Gillian Welch	Revival 1996
Gallows Pole	Led Zeppelin	Led Zeppelin III 1970
Out On The Tiles	Led Zeppelin	Led Zeppelin III 1970
You Shook Me All Night Long	AC/DC	Back In Black 1980
Paper Wings	Gillian Welch	Revival 1996
Celebration Day	Led Zeppelin	Led Zeppelin III 1970
The National Anthem	Radiohead	Kid A 2000
What Do You Do For Money Honey	AC/DC	Back In Black 1980
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Shake A Leg	AC/DC	Back In Black 1980
One More Dollar	Gillian Welch	Revival 1996

(Set 1) Track	Artist	Album Year
Get It While You Can	Janis Joplin	Pearl 1971
Me & Bobby McGee	Janis Joplin	Pearl 1971
A Woman Left Lonely	Janis Joplin	Pearl 1971
Paper Wings	Gillian Welch	Revival 1996
Only One And Only	Gillian Welch	Revival 1996
Honey Baby	Damien Jurado	Rehearsals For Departure 1999
Cry Baby	Janis Joplin	Pearl 1971
Kid A	Radiohead	Kid A 2000
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
One More Dollar	Gillian Welch	Revival 1996
Love The Same	Damien Jurado	Rehearsals For Departure 1999
In Limbo	Radiohead	Kid A 2000
The Ending	Bobby Bare Jr.	Young Criminal's Starvation League 2002
You Shook Me All Night Long	AC/DC	Back In Black 1980
Shoot To Thrill	AC/DC	Back In Black 1980
Treefingers	Radiohead	Kid A 2000
Ohio	Damien Jurado	Rehearsals For Departure 1999
My Baby	Janis Joplin	Pearl 1971
Hells Bells	AC/DC	Back In Black 1980
Friends	Led Zeppelin	Led Zeppelin III 1970

(Set 1) Track	Artist	Album Year
I'll Be Around	Bobby Bare Jr.	Young Criminal's Starvation League 2002
A Woman Left Lonely	Janis Joplin	Pearl 1971
Barroom Girls	Gillian Welch	Revival 1996
Honey Baby	Damien Jurado	Rehearsals For Departure 1999
Orphan Girl	Gillian Welch	Revival 1996
Buried Alive In The Blues	Janis Joplin	Pearl 1971
Friends	Led Zeppelin	Led Zeppelin III 1970
Dig Down	Bobby Bare Jr.	Young Criminal's Starvation League 2002
If You're Feeling Sinister	Belle & Sebastian	If You're Feeling Sinister 1996
Idioteque	Radiohead	Kid A 2000
Saturday	Damien Jurado	Rehearsals For Departure 1999
The National Anthem	Radiohead	Kid A 2000
Hats Off To (Roy) Harper	Led Zeppelin	Led Zeppelin III 1970
Let Me Put My Love Into You	AC/DC	Back In Black 1980
Get It While You Can	Janis Joplin	Pearl 1971
Motion Picture Soundtrack [Acoustic]	Radiohead	Kid A 2000
Tornado	Damien Jurado	Rehearsals For Departure 1999
Trust Me	Janis Joplin	Pearl 1971
Treefingers	Radiohead	Kid A 2000
Acony Bell	Gillian Welch	Revival 1996

(Set 1) Track	Artist	Album Year
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
I'll Be Around	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Honey Baby	Damien Jurado	Rehearsals For Departure 1999
That's The Way	Led Zeppelin	Led Zeppelin III 1970
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister 1996
What Do You Do For Money Honey	AC/DC	Back In Black 1980
Back In Black	AC/DC	Back In Black 1980
You Shook Me All Night Long	AC/DC	Back In Black 1980
Saturday	Damien Jurado	Rehearsals For Departure 1999
Curbside	Damien Jurado	Rehearsals For Departure 1999
Motion Picture Soundtrack [Acoustic]	Radiohead	Kid A 2000
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
Since I've Been Loving You	Led Zeppelin	Led Zeppelin III 1970
Idioteque	Radiohead	Kid A 2000
How To Disappear Completely	Radiohead	Kid A 2000
Shake A Leg	AC/DC	Back In Black 1980
If You're Feeling Sinister	Belle & Sebastian	If You're Feeling Sinister 1996

(Set 1) Track	Artist	Album Year
The National Anthem	Radiohead	Kid A 2000
Idioteque	Radiohead	Kid A 2000
Love The Same	Damien Jurado	Rehearsals For Departure 1999
In Limbo	Radiohead	Kid A 2000
Shoot To Thrill	AC/DC	Back In Black 1980
One More Dollar	Gillian Welch	Revival 1996
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
Tangerine	Led Zeppelin	Led Zeppelin III 1970
Paper Wings	Gillian Welch	Revival 1996
Mehan	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Out On The Tiles	Led Zeppelin	Led Zeppelin III 1970
What Do You Do For Money Honey	AC/DC	Back In Black 1980
Only One And Only	Gillian Welch	Revival 1996
Let Me Put My Love Into You	AC/DC	Back In Black 1980
Motion Picture Soundtrack [Acoustic]	Radiohead	Kid A 2000
Trust Me	Janis Joplin	Pearl 1971
Dig Down	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The Ending	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
Buried Alive In The Blues	Janis Joplin	Pearl 1971

(Set 1) Track	Artist	Album Year
Treefingers	Radiohead	Kid A 2000
Annabelle	Gillian Welch	Revival 1996
Dig Down	Bobby Bare Jr.	Young Criminal's Starvation League 2002
You Shook Me All Night Long	AC/DC	Back In Black 1980
Me & Bobby McGee	Janis Joplin	Pearl 1971
The National Anthem	Radiohead	Kid A 2000
Orphan Girl	Gillian Welch	Revival 1996
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
Idioteque	Radiohead	Kid A 2000
What Difference Does It Make	Bobby Bare Jr.	Young Criminal's Starvation League 2002
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
Since I've Been Loving You	Led Zeppelin	Led Zeppelin III 1970
Given The Dog A Bone	AC/DC	Back In Black 1980
Motion Picture Soundtrack [Acoustic]	Radiohead	Kid A 2000
Hells Bells	AC/DC	Back In Black 1980
Celebration Day	Led Zeppelin	Led Zeppelin III 1970
The Ending	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Pass You By	Gillian Welch	Revival 1996
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002

(Set 1) Track	Artist	Album Year
I'll Be Around	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Have A Drink On Me	AC/DC	Back In Black 1980
How To Disappear Completely	Radiohead	Kid A 2000
The Monk At The Disco	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Tragedy	Damien Jurado	Rehearsals For Departure 1999
Treefingers	Radiohead	Kid A 2000
Get It While You Can	Janis Joplin	Pearl 1971
You Shook Me All Night Long	AC/DC	Back In Black 1980
Me And The Major	Belle & Sebastian	If You're Feeling Sinister 1996
Half Moon	Janis Joplin	Pearl 1971
Acony Bell	Gillian Welch	Revival 1996
One More Dollar	Gillian Welch	Revival 1996
Ohio	Damien Jurado	Rehearsals For Departure 1999
Bullet Through My Teeth	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Tornado	Damien Jurado	Rehearsals For Departure 1999
Move Over	Janis Joplin	Pearl 1971
Pass You By	Gillian Welch	Revival 1996
Kid A	Radiohead	Kid A 2000
Trust Me	Janis Joplin	Pearl 1971
Cry Baby	Janis Joplin	Pearl 1971

(Set 1) Track	Artist	Album Year
Mayfly	Belle & Sebastian	If You're Feeling Sinister 1996
Only One And Only	Gillian Welch	Revival 1996
A Woman Left Lonely	Janis Joplin	Pearl 1971
My Baby	Janis Joplin	Pearl 1971
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
By The Mark	Gillian Welch	Revival 1996
Saturday	Damien Jurado	Rehearsals For Departure 1999
Everything In It's Right Place	Radiohead	Kid A 2000
Motion Picture Soundtrack [Acoustic]	Radiohead	Kid A 2000
Tornado	Damien Jurado	Rehearsals For Departure 1999
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
Immigrant Song	Led Zeppelin	Led Zeppelin III 1970
Given The Dog A Bone	AC/DC	Back In Black 1980
The Monk At The Disco	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Treefingers	Radiohead	Kid A 2000
Ohio	Damien Jurado	Rehearsals For Departure 1999
Like Dylan In The Movies	Belle & Sebastian	If You're Feeling Sinister 1996
Trust Me	Janis Joplin	Pearl 1971
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Annabelle	Gillian Welch	Revival 1996

(Set 1) Track	Artist	Album Year
Back In Black	AC/DC	Back In Black 1980
That's The Way	Led Zeppelin	Led Zeppelin III 1970
How To Disappear Completely	Radiohead	Kid A 2000
Acony Bell	Gillian Welch	Revival 1996
Love The Same	Damien Jurado	Rehearsals For Departure 1999
Given The Dog A Bone	AC/DC	Back In Black 1980
Dig Down	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Rock And Roll Ain't Noise Pollution	AC/DC	Back In Black 1980
I'll Be Around [Reprise]	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Treefingers	Radiohead	Kid A 2000
Barroom Girls	Gillian Welch	Revival 1996
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister 1996
Me & Bobby McGee	Janis Joplin	Pearl 1971
Curbside	Damien Jurado	Rehearsals For Departure 1999
Pass You By	Gillian Welch	Revival 1996
Friends	Led Zeppelin	Led Zeppelin III 1970
What Do You Do For Money Honey	AC/DC	Back In Black 1980
Eyes For Windows	Damien Jurado	Rehearsals For Departure 1999
Stay In Texas	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Trust Me	Janis Joplin	Pearl 1971

(Set 1) Track	Artist	Album Year
Half Moon	Janis Joplin	Pearl 1971
If You're Feeling Sinister	Belle & Sebastian	If You're Feeling Sinister 1996
Mehan	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Tangerine	Led Zeppelin	Led Zeppelin III 1970
Saturday	Damien Jurado	Rehearsals For Departure 1999
The Monk At The Disco	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Mercedes Benz	Janis Joplin	Pearl 1971
Gallows Pole	Led Zeppelin	Led Zeppelin III 1970
Ohio	Damien Jurado	Rehearsals For Departure 1999
The Ending	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Eyes For Windows	Damien Jurado	Rehearsals For Departure 1999
Back In Black	AC/DC	Back In Black 1980
Everything In It's Right Place	Radiohead	Kid A 2000
Hells Bells	AC/DC	Back In Black 1980
That's The Way	Led Zeppelin	Led Zeppelin III 1970
Tornado	Damien Jurado	Rehearsals For Departure 1999
By The Mark	Gillian Welch	Revival 1996
I'll Be Around [Reprise]	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Bron-Yr-Aur Stomp	Led Zeppelin	Led Zeppelin III 1970
Celebration Day	Led Zeppelin	Led Zeppelin III 1970

(Set 1) Track	Artist	Album Year
Trust Me	Janis Joplin	Pearl 1971
The Ending	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Kid A	Radiohead	Kid A 2000
Love The Same	Damien Jurado	Rehearsals For Departure 1999
Out On The Tiles	Led Zeppelin	Led Zeppelin III 1970
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
Barroom Girls	Gillian Welch	Revival 1996
Idioteque	Radiohead	Kid A 2000
Bron-Yr-Aur Stomp	Led Zeppelin	Led Zeppelin III 1970
The Stars Of Track And Field	Belle & Sebastian	If You're Feeling Sinister 1996
Mayfly	Belle & Sebastian	If You're Feeling Sinister 1996
Annabelle	Gillian Welch	Revival 1996
If You're Feeling Sinister	Belle & Sebastian	If You're Feeling Sinister 1996
Ohio	Damien Jurado	Rehearsals For Departure 1999
Tangerine	Led Zeppelin	Led Zeppelin III 1970
Saturday	Damien Jurado	Rehearsals For Departure 1999
Pass You By	Gillian Welch	Revival 1996
Me & Bobby McGee	Janis Joplin	Pearl 1971
How To Disappear Completely	Radiohead	Kid A 2000
Given The Dog A Bone	AC/DC	Back In Black 1980

(Set 1) Track	Artist	Album Year
If You're Feeling Sinister	Belle & Sebastian	If You're Feeling Sinister 1996
Have A Drink On Me	AC/DC	Back In Black 1980
Me & Bobby McGee	Janis Joplin	Pearl 1971
Hats Off To (Roy) Harper	Led Zeppelin	Led Zeppelin III 1970
Morning Bell	Radiohead	Kid A 2000
Shoot To Thrill	AC/DC	Back In Black 1980
Everything In It's Right Place	Radiohead	Kid A 2000
What Difference Does It Make	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Curbside	Damien Jurado	Rehearsals For Departure 1999
The Fox In The Snow	Belle & Sebastian	If You're Feeling Sinister 1996
Back In Black	AC/DC	Back In Black 1980
Treefingers	Radiohead	Kid A 2000
Let Me Put My Love Into You	AC/DC	Back In Black 1980
My Baby	Janis Joplin	Pearl 1971
Pass You By	Gillian Welch	Revival 1996
Tangerine	Led Zeppelin	Led Zeppelin III 1970
Annabelle	Gillian Welch	Revival 1996
Rehearsals For Departure	Damien Jurado	Rehearsals For Departure 1999
Ohio	Damien Jurado	Rehearsals For Departure 1999
Get Me Away From Here, I'm Dying	Belle & Sebastian	If You're Feeling Sinister 1996

(Set 1) Track	Artist	Album Year
Buried Alive In The Blues	Janis Joplin	Pearl 1971
Barroom Girls	Gillian Welch	Revival 1996
Tragedy	Damien Jurado	Rehearsals For Departure 1999
Half Moon	Janis Joplin	Pearl 1971
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister 1996
Mayfly	Belle & Sebastian	If You're Feeling Sinister 1996
I'll Be Around	Bobby Bare Jr.	Young Criminal's Starvation League 2002
Tangerine	Led Zeppelin	Led Zeppelin III 1970
A Woman Left Lonely	Janis Joplin	Pearl 1971
Rock And Roll Ain't Noise Pollution	AC/DC	Back In Black 1980
Motion Picture Soundtrack [Acoustic]	Radiohead	Kid A 2000
In Limbo	Radiohead	Kid A 2000
Only One And Only	Gillian Welch	Revival 1996
Rehearsals For Departure	Damien Jurado	Rehearsals For Departure 1999
Move Over	Janis Joplin	Pearl 1971
Seeing Other People	Belle & Sebastian	If You're Feeling Sinister 1996
Ohio	Damien Jurado	Rehearsals For Departure 1999
Letters & Drawings	Damien Jurado	Rehearsals For Departure 1999
Have A Drink On Me	AC/DC	Back In Black 1980
Pass You By	Gillian Welch	Revival 1996

(Set 1) Track	Artist	Album	Year
Tragedy	Damien Jurado	Rehearsals For Departure	1999
The National Anthem	Radiohead	Kid A	2000
Let Me Put My Love Into You	AC/DC	Back In Black	1980
The Boy Done Wrong Again	Belle & Sebastian	If You're Feeling Sinister	1996
Judy And The Dream Of Horses	Belle & Sebastian	If You're Feeling Sinister	1996
Honey Baby	Damien Jurado	Rehearsals For Departure	1999
Hats Off To (Roy) Harper	Led Zeppelin	Led Zeppelin III	1970
Buried Alive In The Blues	Janis Joplin	Pearl	1971
Treefingers	Radiohead	Kid A	2000
Annabelle	Gillian Welch	Revival	1996
Only One And Only	Gillian Welch	Revival	1996
Get Me Away From Here, I'm Dying	Belle & Sebastian	If You're Feeling Sinister	1996
Barroom Girls	Gillian Welch	Revival	1996
Tangerine	Led Zeppelin	Led Zeppelin III	1970
Curbside	Damien Jurado	Rehearsals For Departure	1999
Bron-Yr-Aur Stomp	Led Zeppelin	Led Zeppelin III	1970
In Limbo	Radiohead	Kid A	2000
Me And The Major	Belle & Sebastian	If You're Feeling Sinister	1996
Cry Baby	Janis Joplin	Pearl	1971
Have A Drink On Me	AC/DC	Back In Black	1980

Randomness Homework

To follow up from today's iPod Shuffle Activity and develop more ideas about randomness and human's intuitions about randomness, read the following blog and answer the following questions based on the reading.

Blog: *Science in Action*
http://sxxz.blogspot.com/2005/08/what-is-randomness.html

Questions:

1. How should we define randomness?

2. Why is it hard to detect randomness?

3. Describe two ways why randomness is important.

4. Traditionally, shuffling cards does not truly randomize a deck of cards. What would it take in order to get a deck of cards randomized well enough for practical purposes?
5. When flipping a fair coin for a large number of random flips, what is the percentage of times you would expect to see five heads in a row?

Randomness Homework
Solutions

Questions:

1. How should we define randomness?

   We should define randomness by what it isn't.

2. Why is it hard to detect randomness?

   Common sense makes it hard for humans to detect randomness. We often see patterns and structure in sequences, even when they are truly random and no pattern exists. In addition, we tend to assume non-

randomness when we see repetitions in a sequence.

3. Describe two ways why randomness is important.

   Internet Security – Cryptographic systems depend on random numbers.

   Gambling – Casinos depend on the gambler's misconceptions about randomness. For example, gamblers often think they can predict the behavior of slot machines. Lotteries also rely on randomness.

4. Traditionally, shuffling cards does not truly randomize a deck of cards. What would it take in order to get a deck of cards randomized well enough for practical purposes?

   To get a deck of cards randomized enough for practical purposes requires more than seven imperfect shuffles.

5. When flipping a fair coin for a large number of random flips, what is the percentage of times you would expect to see five heads in a row?

   3% of the time

Overview of Lesson 2

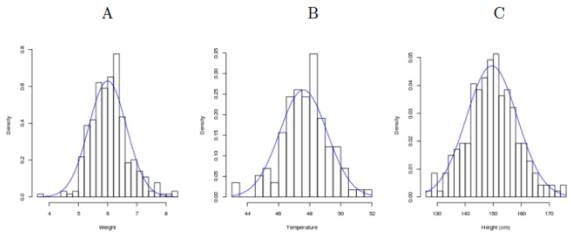
Topic of Lesson: Normal Distributions		**Date:**
Summary of Lesson: This lesson centers on the normal distribution, why it is important, and when it is appropriate to use it. A brief review of the normal distribution including the basic properties of the normal distribution will be presented first. Students will learn that if a distribution is normal, then the mean and standard deviation can tell us everything else about the distribution. The normal distribution allows us to find probabilities (areas under a curve) and students will solve two related problems: the unknown percentage problem and the unknown value problem. Furthermore, students will explore two datasets (song lengths and album prices) and recognize under which dataset the normal model is more appropriate.		
Prerequisites: • 68-95-99.7 Rule (Empirical Rule) • Recognizing Skewed Distribution	**Activities:** N/A	**Estimated Time:** Approximately 75 minutes
Learning Objectives ***Students will be able to …*** • recognize when a normal distribution/model is appropriate to make statistical inferences. • use the empirical rule and normal models to estimate the percentage of observations falling within 1, 2, or 3 standard deviations of the mean. • know how to find the percentage of observations falling below any value in a normal model (i.e., unknown percentage problem). • know how to find the value(s) of a normal model given a percentage below, above, or between (i.e., unknown value problem).	**Assessments** • Optional HW assignment with problems of your choice (see *normal_dist_hw_sample.docx* file for possible homework problems)	
CCSSM Standards Addressed • *S.ID.4.* Use the mean and standard deviation of a data set to fit a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not	**CCSSM Mathematical Practices Emphasized** • *MP1.* Make sense of problems and persevere in solving them. • *MP4.* Model with mathematics. • *MP5.* Use appropriate tools strategically.	

appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.	

Teaching Materials

- Varies (your decision) – graphing calculators, computers, tables, handouts, etc.
- *SongLengths_AlbumPrices.xlsx* file

Learning Plan for Lesson 2

Teacher Activity	Estimated Time
1. *Normal Distribution Review (if needed)* **Review the basic properties of the normal distribution. For example:** **a) Provide visual examples of normal distributions and ask students questions.** *A) Weights (g) of newly born rat pups* *B) Mean annual temperatures (°F) in Ann Arbor, Michigan* *C) Heights (cm) of 14-year-old boys in Oxford, England*  *Question:* *In what ways are these distributions similar and different?*	~15 min (longer if more review is needed)

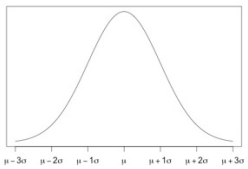
Possible Answer:
Means and spreads all differ, but the shapes are quite similar (sort of bell-shaped).

b) Provide the definition of the normal distribution
Example: Many distributions we look at have a shape similar to those above. Most of the data lies close to the mean, and the left and right sides are symmetric. We call it "bell-shaped" and the best example is called the "Normal" distribution, which is important in statistics.

c) (OPTIONAL) Discuss the notation used to represent normal distributions
Example: Normal distributions all have the same shape and they differ only in mean (μ) and spread (σ). We write $X \sim N(\mu, \sigma)$ to say that random variable X is normally distributed with center μ and spread (officially: standard deviation) σ.

d) (OPTIONAL) Discuss the standard normal distribution and how to "standardize"
Example: The standard normal distribution has center $\mu = 0$ and spread $\sigma = 1$. We can "standardize" any normal distribution to make it have center 0 ($\mu = 0$) and spread 1 ($\sigma = 1$) by subtracting μ and dividing by σ. If X has a $N(\mu, \sigma)$ distribution, then $z = \frac{x-\mu}{\sigma}$ has a $N(0, 1)$ distribution. We use the standardized versions to say how many standard deviations (σ) an observation is from the mean (μ). To "unstandardized" you can solve the z-score formula for x and get $x = \mu + z\sigma$.

e) Sketching normal curves
Example: To sketch a normal curve, start by drawing a smooth curve opening downward. Although this curve extends in either direction to infinity, you need only sketch the first three standard deviations. After three standard deviations away from the mean, there is so little left of the curve that it is not worth sketching. The place where the bell shape changes from curving downward to curving back up is called the inflection point and occurs exactly one standard deviation away from the mean.

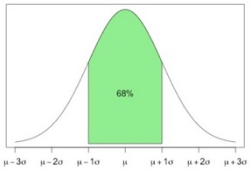


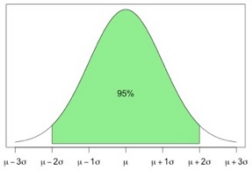
Example: Sketch $N(1, 2)$.

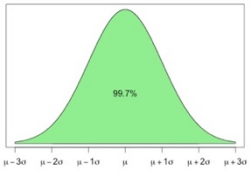
f) Provide a definition of normal probabilities and discuss the Empirical Rule (i.e., 68-95-99.7% Rule)

Normal probabilities are areas under the normal curve and the total area under a normal curve is 1. In a normal model we have three facts that are summarized in a rule called the Empirical Rule (or 68-95-99.7 Rule):

- *approximately 68% of the values fall within 1 standard deviation of the mean*
- *approximately 95% of the values fall within 2 standard deviations of the mean*
- *approximately 99.7% of the values fall within 3 standard deviation of the mean*



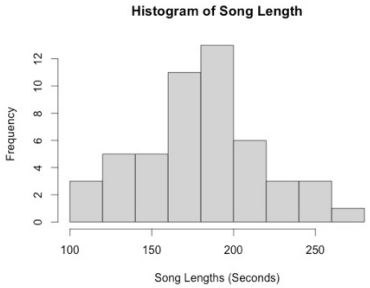


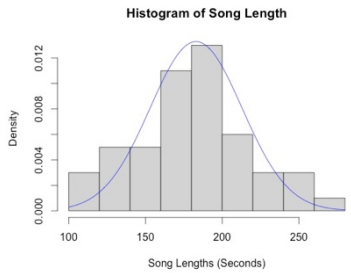


2. *Using the Normal Distribution – Song Lengths Dataset*

Part 1: Explore the Song Lengths Dataset

50 randomly selected songs from iTunes were selected and the lengths of the songs (seconds) were recorded (*SongLengths_AlbumPrices.xlsx* file). The song lengths follow a normal distribution and summary statistics are show below.



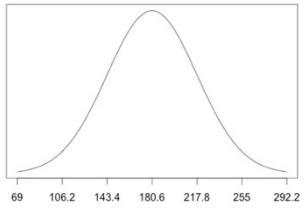


Minimum	103.0
Q1	160.0
Median	181.5
Q3	201.0
Maximum	261.0
Mean	180.6
Std. Dev.	37.20

~ 30 min

Part 2: Use the song lengths normal model to find various probabilities

Because song lengths can be represented by a normal distribution, we can calculate various unknown percentage and unknown value problems. Below is a sketch of this normal model, along with examples of various unknown percentage problems and unknown value problems that can be worked through at your discretion.



Examples of Unknown Percentage Problems

These calculations can be done a variety of ways: by hand (e.g., using tables) or with technology (e.g., calculator, GeoGebra). Answers are provided in parentheses.

1. What percentage of song lengths is (or "What is the probability that a song length is")

   a. between 143.4 and 217.8 seconds? (68%)
   b. between 69 and 292.2 seconds? (99.7%)
   c. above 180.6 seconds? (50%)
   d. above 106.2 seconds? (97.5%)
   e. below 217.8 seconds? (84%)
   f. below 143.4 seconds? (16%)

   g. between 150 and 180 seconds? (28.8%)
   h. between 80 and 200 seconds? (69.6%)
   i. above 155 seconds? (75.4%)
   j. above 280 seconds? (0.38%)
   k. below 190 seconds? (60%)
   l. below 140 seconds? (13.8%)

Examples of Unknown Value Problems

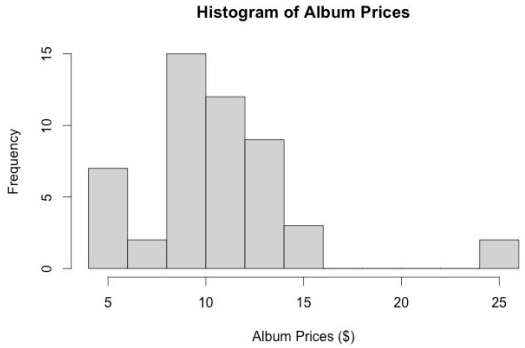
These calculations can be done a variety of ways: by hand (e.g., using tables) or with technology (e.g., calculator, GeoGebra). Answers are provided in parentheses.

1. Find the song length that cuts off the largest 2.5% of the values. (253.51 seconds)
2. What song length separates the highest 25% of song lengths from the rest? (206.67 seconds)
3. What song length separates the lowest 25% of song lengths from the rest? (155.53 seconds)

4. Find the song length, which has 57% of the song lengths below it (i.e., 57th percentile). Before doing this, write down whether this will be more or less than 180.6 seconds and explain why. (More than 180.6 seconds. 187.1 seconds)

3. *Skewed Distributions – Album Prices Dataset*

Part 1: Explore the Album Prices Dataset
50 albums were randomly selected from iTunes and the prices (dollars) of each album were recorded (*SongLengths_AlbumPrices.xlsx* file). Here, we note that the album prices do **NOT** follow a normal distribution, but rather represent a skewed right distribution. Below is a histogram of the album prices and summary statistics.



Minimum	4.95
Q1	9.99
Median	10.99
Q3	12.99
Maximum	24.99
Mean	11.19
Std. Dev.	3.92

Part 2: Overall question to explore/discuss
Can the procedures we used above with song lengths be used all of the time? When is it appropriate to use those procedures?
No. We need to make sure our data represents a normal model before applying the procedures we used above.

Discuss this idea at your discretion.:

Example: Can you find the percentage of album prices that are between, say, $7.27 and $15.11, using the same techniques as before In other words, is it true that because 7.27 and 15.11 are exactly 1 standard deviation away from the mean (11.19), the percentage of albums priced between $7.27 and $15.11 is 68%?

No, it is not appropriate to use a normal model given our dataset. In the example above, the correct percentage of album prices between $7.27 and $15.11 is 78% (39/50), not the 68% that would be calculated using the Empirical Rule. It's important to first check the histogram of a dataset to determine when a normal model is appropriate. Students would have to use a different

~20 min

technique to calculate this percentage (e.g., count the number of albums that are between $7.27 and $15.11 and divide by the total sample size (50)).	
4. *Wrap Up* **Take-Home Messages** ➢ We used the mean and standard deviation of the song lengths dataset to fit a normal distribution and to estimate various population percentages and find unknown song lengths. However, there are datasets for which such procedures are not appropriate. ➢ It is important the plot the data and check the distribution before using any procedures to calculate various population percentages. Note: A histogram is the most appropriate way to check the distribution of a continuous variable. ➢ Normal probabilities are areas under a curve. ➢ We can use the Empirical Rule (68-95-99.7 Rule) with a normal distribution.	~ 5 min
5. *Homework* (OPTIONAL) At your discretion, assign homework problems of your choice. For example, you could assign problems where students need to use the mean and standard deviation of a dataset to fit a normal distribution and to estimate population percentages. Or assign problems in which students are able to recognize that these procedures are not appropriate. The file, *normal_dist_hw_sample.docx*, contains sample problems and the solutions can be found in the file, *normal_dist_hw_sample_solutions.docx*.	

Normal Distribution Homework
Sample Problems

1. ACT math scores are known to have mean 21 with standard deviation 5. For each of the following, draw a picture of the scenario and calculate the corresponding percentage.

   a. The percentage of ACT math scores between 11 and 31.

   b. The percentage of ACT math scores less than 16.

   c. The percentage of ACT math scores greater than 36.

   d. Find the ACT math score, which has 84% of scores below it (i.e., the 84th percentile). Before finding the value, write down whether this score will be more or less than 21 and explain why.

2. Tests of intelligence are designed to return IQs having a normal distribution centered at 100 with standard deviation 15.

   a. Use the "Empirical Rule" to make three statements about IQ tests. For each say what percentage of IQs falls within a specific range about 100 (give the percentage and the range).

   b. Using your answer from part (a), what percentage of IQs are greater than 145? Explain how you calculated this number.

   c. For each of the following, draw a picture of the scenario and calculate the corresponding proportion.

      i. The proportion of IQs over 140.

      ii. The proportion of IQs between 90 and 110.

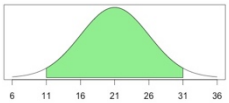
      iii. The proportion of IQs below 122.

      iv. Find the IQ number, which has 37% of IQs below it (i.e., the 37th percentile). Before finding the value, write down whether this number will be more or less than 100 and explain why.

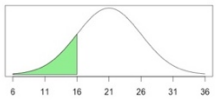
Normal Distribution Homework
SOLUTIONS

1. ACT math scores are known to have mean 21 with standard deviation 5. For each of the following, draw a picture of the scenario and calculate the corresponding percentage.

   a. The percentage of ACT math scores between 11 and 31. *(95%)*



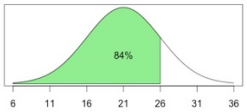
b. The percentage of ACT math scores less than 16. *(16%)*



c. The percentage of ACT math scores greater than 36. *(0.15%)*



d. Find the ACT math score, which has 84% of scores below it (i.e., the 84th percentile). Before finding the value, write down whether this score will be more or less than 21 and explain why. *(More than 21 because 84% > 50% and the mean of the distribution is 21. ACT math score is 26.)*



2. Tests of intelligence are designed to return IQs having a normal distribution centered at 100 with standard deviation 15.

a. Use the "empirical rule" to make three statements about IQ tests. For each say what percentage of IQ's falls within a specific range about 100 (give the percentage and the range).

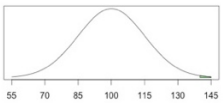
- *68% of the data falls within 85 to 115.*
- *95% of the data falls within 70 to 130.*
- *99.7% of the data falls within 55 to 145.*

b. Using your answer from part (a), what percentage of IQs are greater than 145? Explain how you get the number.

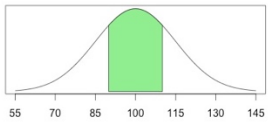
145 is three standard deviations to the right of the mean. Therefore, (100% - 99.7%)/2 = 0.15% of the data lies above 145.

c. For each of the following, draw a picture of the scenario and calculate the corresponding proportion.

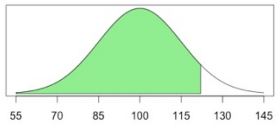
i. The proportion of IQs over 140. *(0.0038)*



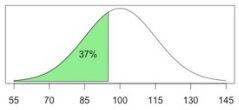
ii. The proportion of IQs between 90 and 110. *(0.495)*



iii. The proportion of IQs below 122. *(0.9288)*



iv. Find the IQ number, which has 37% of IQs below it (i.e., 37th percentile). Before finding the value, write down whether this will be more or less than 100 and explain why. *(Less than 100 because 37% < 50% and the mean is at 100. IQ number is 95.02)*



Song Lengths and Album Prices Dataset

Song Lengths	
173	261
130	176
132	202
135	215
185	218
163	174
147	103
249	194
195	168
181	154
167	125
182	180
149	175
218	156
231	183
197	119
144	187
139	103
221	198
191	193
188	209
159	231
214	183
246	170
246	163

Album Prices	
9.99	9.99
12.99	9.99
12.99	4.95
11.99	10.99
24.99	12.99
12.99	5.99
12.99	7.99
10.99	9.99
5.99	9.99
7.99	11.99
9.99	9.99
11.99	12.99
9.99	12.99
9.99	4.99
11.99	9.99
15.99	4.99
11.99	9.99
9.99	15.99
9.99	5.99
12.99	9.99
14.99	9.99
11.99	11.99
11.99	24.99
11.99	4.99
12.99	11.99

Overview of Lesson 3

Topic of Lesson:		**Date:**
Observational Studies and Experiments		
Summary of Lesson: This lesson centers on observational studies, experiments, and the advantages of a randomized experiment. A very brief review of the difference between categorical and quantitative variables should be presented if needed. The remainder of the lesson is an activity that first guides students through the definitions of an observational study and an experiment and then uses a web applet to help students learn the advantages of a randomized experiment. The main experiment proposed focuses on investigating the effectiveness of studying with music using a treatment and control group. The concepts of lurking variables and how randomization (i.e., randomly assigning treatments) "evens out" the lurking variable are also explored throughout this activity. *Note: This web applet works on computers, tablets, and smartphones.*		
Prerequisites: • Categorical vs. Quantitative Variable • Calculating a mean	**Activities:** Does Music Help Us Study? – Part 1	**Estimated Time:** Approximately 100 minutes (depends on the amount of time you want to spend on the activity)
Learning Objectives ***Students will be able to …*** • learn the difference between observational studies and experiments. • recognize the advantages of a randomized experiment. • learn what a lurking variable is and provide examples. • understand the advantage of a randomized experiment by learning that randomly assigning treatments tends to balance out lurking variables so that we can say the differences we observed are caused by the explanatory variable (this is called causal inference).		**Assessments** • Optional HW assignment with problems of your choice (see *obs_exp_hw_sample.docx* file for possible homework problems)

CCSSM Standards Addressed	CCSSM Mathematical Practices Emphasized
• *S.IC.3.* Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.	• *MP1.* Make sense of problems and persevere in solving them. • *MP3.* Construct viable arguments and critique the reasoning of others. • *MP5.* Use appropriate tools strategically.

Teaching Materials

- Copies of the *Does Music Help Us Study? – Part 1* activity (one per group or one per student)
- Copies of *Additional Problems - Does Music Help Us Study? – Part 1* document if you choose to use these extra problems (one per group or one per student)
- Access to a computer, tablet, or smartphone
- Completed sets (one per group or one per student) of cards for #5 on the activity OR blank sets of cards that students can write information on themselves

Learning Plan for Lesson 3

Teacher Activity	Estimated Time
1. *Categorical vs. Quantitative Variable Review (if needed)* **Review the definitions of categorical and quantitative variables** *There are two main types of data:* *a) Quantitative – takes numeric values we can average* *b) Categorical – falls into one or more categories. The categories can have numeric labels (like zip codes), but it makes no sense to average them. (Some call this "Qualitative.")* **Examples** Identify each of the following. Answers are provided in parentheses. a) The cell phone carrier you use (*Categorical*) b) The amount you spend on lunch this week (*Quantitative*)	~ 5 min
2. *Does Music Help Us Study? – Part 1 Activity*	

Pass out the activity. Access to a computer, tablet, or smart phone will be needed to use the web applet mentioned in the activity. Note that use of this web applet does not occur until Problem #6 in the activity. Problems 1 – 5 do not require the use of the web applet.

It's recommended to let students work through this activity together in groups of 2 – 3 where each student or each group uses a computer/tablet/smartphone. Groups of 2 using 1 computer (and working together) have worked out well in the past. If you have each student use a computer/tablet/smartphone, note that they will get slightly different answers than others when using the web applet. This may frustrate those working together, but the web applet generates data anytime it is used which results in some variability.

The activity also works best when students discuss and think about their answers together as opposed to solely completing the activity without any discussion.

Problems 1 - 3

- #1 introduces an observational study and #2 introduces an experiment.
- In #3, you are asked to identify which study was an observational study and which was an experiment.

Letting students work through these questions as a group and then having a large group discussion or individually checking in with each group works best. For example, when asking for a response in #1a, students may struggle to think of a response to use and having a group discussion can sometimes help with this. It's also good to hear a bunch of ways the response can be measured – there isn't one correct answer.

Problems 4 – 5

- These questions work through a randomized experiment and guide students through the advantages of randomly assigning treatments.
- #4 describes a lurking variable and then asks to identify possible lurking variables (*the ones used in this experiment are IQ Score (quantitative) and whether or not a student took an SAT Prep course (categorical)).*
- #5 begins the "simulation" part of this activity, but by hand. Students need 12 cards (one for each participant in the experiment) that have the name, IQ score and whether or not an SAT Prep course was taken written on it. *To save time during this activity, it is recommended to have these cards completed and ready beforehand. However, you may also choose to have students complete these cards themselves. These cards can be created using any time of paper.*

~ 90 min

(Estimated time depends on how the activity is completed. For example, working through the problems and applet without any difficulties or discussions can take about 50 minutes. Discussing the problems and troubleshooting any difficulties adds additional time.)

#5 asks students to do the randomization "by hand." After students make their two piles and record their observations, make sure the students leave their two piles (treatment and control) as is because you may want your students to come back to these cards at the end of the activity when they work on #13. Question #13 requires students to look at the categorical lurking variable (whether or not an SAT Prep course was taken) and there are additional questions (if you choose to use them) that can be asked in which students will need to record what proportion in each group took the SAT prep course. These additional questions can be found below under Problem 13.

Graphical addition & a great question to ask regarding #5

In #5a, students are asked to compute the mean for each group and take the difference. If you have time, you can have students report the difference to you and it is recommended to create a randomization plot with the students' differences. It is suggested to create a dot plot of the differences and discuss the plot as a group. This gives a nice introduction to what will be seen using the web applet (#6 – 9), but on a much smaller scale. For example, GeoGebra (www.geogebra.org) can be quickly used to make a dot plot. Once you at at the website, you would click on "Start GeoGebra", select "Spreadsheet", enter the differences in column A, select column A, click on the purple histogram at the top of the window and select "One Variable Analysis." A histogram will automatically appear and if you click on the double arrows after Histogram at the top of the plot, you will see the option for a "Dot Plot." Additional instructions can be found using a quick Google search.

By looking at just the class data, students can better understand how the randomization plot is created. A question we often ask students about randomization plots is **"How is one dot created?"** ***This is a very hard question for students and they often struggle with it.*** For example, sample student responses have been "It's the difference in sample sizes." or "It's a data point." The answer, in the context of what the students did by hand, is "12 cards were made with an IQ score recorded on each one. The cards were shuffled and randomly placed into two groups (each of size 6) with one group labeled Treatment and the other labeled Control. The location of the dot gives the difference in the mean IQ Score in the Treatment group and the mean IQ Score in the Control group (Treatment – Control)." This question is a lot easier to discuss using only a few data points rather than thousands that can be created with the applet.

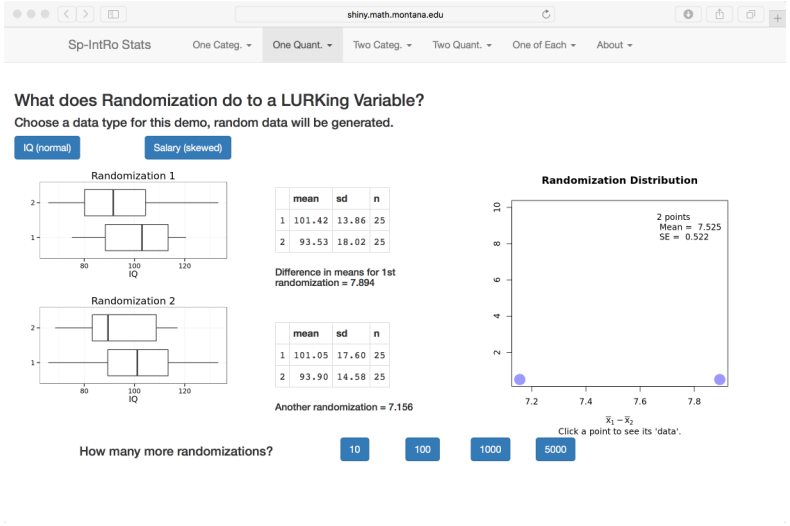
If you ask this same question (**"How is one dot created?"**) once they have used the web applet (where we started with a larger sample size and have more points plotted), the answer is "50 cards were made with an IQ score recorded on each one. The cards were shuffled and randomly placed into two groups (each of size 25) with one group labeled Treatment and the other labeled Control. The location of the dot gives the difference in the mean IQ Score in the Treatment group and the mean IQ Score in the Control group (Treatment – Control)."

Furthermore, asking this type of question helps set up students to make connections later on between what they did by hand and what the web applet is doing. Both are doing the same thing, but the web applet is much faster!

Problems 6 – 9

- Access to a computer, tablet, or smartphone is needed.
- The web applet used for this activity (http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/) is introduced and the directions in #6 help set up the app/demonstration.
- #6 – 9 ask questions regarding the use of the web applet.

#6 starts the use of the web applet. Students should be able to follow the directions and select IQ (normal) and the plot below will appear on their screen.



Note that the applet automatically displays two randomizations. This is similar to having the students shuffle the cards into two groups twice. Students might ask you what these two randomizations mean.

The summary statistics for Randomization 1 and Randomization 2 will likely differ for each person/group using this web applet. For instance, the difference in means for the 2nd randomization above is "Another randomization = 7.156" where, for another group, it could be -3.741 or something different. It's good to point out that there is variability occurring because anything can occur when we "shuffle" and sort participants into groups. You can relate this back to the randomization plot you created with the class data because there were likely different differences recorded and plotted.

There are two dots plotted in the Randomization Distribution plot (far right-hand plot). Each of these dots represents the difference in means for the 1st and 2nd randomization recorded below each summary table. The following questions about these points help guide students to the main advantage of using randomization in an experiment.

1. "What do we want this difference to be?"
Ideally, we want the difference to be 0 because this means that the two groups are similar in IQ score and so any results we make in this experiment can be attributed to the treatment. The lurking variable (IQ score) is being "evened out" between the two groups.

2. "What does a difference of -4.52 mean?"
A difference of -4.52 means that the control group has an higher average IQ score which can affect the results of the study.

You can relate each summary table (which doesn't change as you select 10, 100, 1000, or 5000 more randomizations – it's always the summary statistics from the very first shuffle when IQ (normal) was selected) and corresponding dots back to what the students did by hand when they shuffled the cards manually. In other words, **"How is this web applet mimicking what they did by hand with the cards?"** A discussion relating the "by hand" shuffling to the app can help students understand the benefits of using a computer to do the randomization for you.

When using this applet, students may just start clicking buttons at first (i.e., the 10, 100, 1000, 5000 buttons). If at any point students need to "start over" to get back on track with the activity, they can click IQ (normal) and the plot will reset (with different numbers and proportion) to display just two points.

A cool feature of this app is the ability to click on dots in the randomization distribution and see what that proportion that dot represents. When you click on a point, you will notice that the display for **Randomization 2** (the boxplot and table of proportions) will change. Randomization 1 will always remain the same. This new table displays the information for that particular plot. This is a great feature to use to further explore and address what each dot really represents.

Problems 10 – 12

- These problems wrap up the investigation of the quantitative lurking variable (IQ Score) and guide students to learn the advantages of randomized experiments.

For problems #10 – 12, remind students to refer back to their randomization plot from #7. These are important questions to discuss.

Problem #12 is a very important question (the most important out of #10, 11, & 12) and addresses the main advantage of a randomized experiment. In the randomization plot from #7, you can see that there are values such as -7.5, -4.5. etc. (i.e., values far from 0) which indicate that the mean IQ scores are NOT always the same between the two treatment groups, but ON AVERAGE (indicated by the mean of the randomization plot after thousands of simulations), they are.

Note that some experiments do not randomly assign participants/subjects into treatment/control groups, but rather just assign them to groups. That is okay, but not an ideal study design. The point we want to emphasize is that experiments assign treatments and observational studies do not. An experiment is called a randomized experiment if it randomly assigns treatments.

Problem 13

- This last question requires some reflection on the entire activity and asks students to predict/explain what would occur if we were to investigate another lurking variable – this time a categorical lurking variable (whether or not an SAT Prep course was taken).
  #13 is a similar situation that was just worked throughout the activity where the only difference being in that this question revolves around a categorical (SAT Prep course) lurking variable instead of a quantitative (IQ Score) lurking variable.

Original #13

Students can answer this question as is based on the knowledge gained throughout the activity. The only difference being that this question involves a categorical lurking variable.

"New" #13 plus additions:

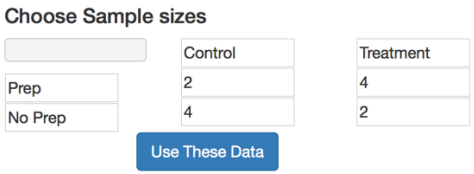
Students can explore what happens with this categorical lurking variable on the same web applet app they just used. The following questions can replace Problem #13 and these questions provide students the opportunity to use the web applet again to explore this concept. **For your convenience, the word document, *L3_activity_additions.docx*, contains these new additions and the corresponding answers can be found in the *L3_activity_additions_solutions.docx* file.**

Extra Questions:

13. Another lurking variable would be the fact that some people have taken a short course as an SAT prep course and others haven't. If the SAT prep course does what it claims, then it could be the reason for one group to score higher than the other. We will look at the proportions who have taken an SAT prep course in the treatment and control groups. We are now looking at proportions instead of means because this lurking variable (took SAT Prep course) is a categorical variable.

a. Compute the proportion of Yes's in the two groups of cards you had shuffled, and subtract. Write the proportions and the difference here.

b. Again go to http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/ and click Lurking Demo, but this time under the One Categ. header. Enter the data as follows:



and click Use These Data. Run 5000 "more simulations" and record the mean and SE of the differences.

c. You will have a few shuffles that give -1 or 1. How could that happen?

→ Students often have a difficult time with this question because they don't fully understand the plot and how it was created. This is a good question to ask because it helps students think more about the plot and how it was created. In other words, "What does one dot represent?"

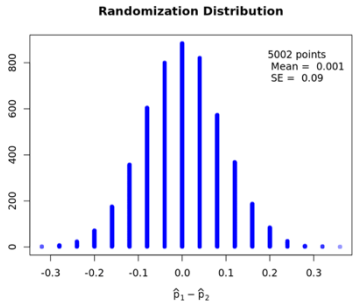
d. The plot gets more interesting with larger counts. Suppose we are randomly assigning 100 people to our two groups, and that 28 of them have taken SAT prep, 72 have not. Of those in the Music group, 10 of them have taken an SAT Prep course and 40 have not. Of those in the No Music group, 18 have taken an SAT Prep course and 32 have not. Enter the data as follows:



and click Use These Data. What proportions of "Prep" are in the treatment and what differences do you get for the first two randomizations?

e. Run 5000 randomizations. Sketch your plot here.

→ This is what the randomization plot should look like. It's always nice to see what happens when we have more data to analyze. You could ask students to compare this randomization plot to the randomization plot they saw in #13b.



f. Compare with other groups. Do the pictures look the same?

→ Questions 14 and 15 below lead the students to answer the original question posed. These questions wrap up what they saw working with a categorical lurking variable. Question 16 asks them to put everything together and is the most important question of the three.

14. When we randomly assign people to two groups:

a. Is it possible for a categorical lurking variable like SAT prep to be imbalanced across the two groups? Explain.
b. Will the lurking SAT prep variable "usually" be poorly balanced across the two groups? Explain?

→ Students may struggle with these questions and have trouble understanding the wording, especially "usually". Remind students to look back at their randomization plot in #13e. Oftentimes, students think that it isn't possible for a lurking variable to be imbalanced across the two groups because they focus on the center of the randomization plot. Remind students what the endpoints of that that plot represent. It is possible that the lurking variable is imbalanced across the two groups, but on

average, this is not the case. Part a) of this question is asking students to consider one point, while part b) is asking them to consider the average of thousands of points. 15. In general, how similar is the proportion of people who have taken SAT prep in the treatment group to the same proportion in the control group? → This question can seem like a "repeat" for some students and it is repeating what was emphasized in #14. However, it has been found that multiple questions that are "repetitive" in nature, help students take-away the main concept of this activity – randomization tends to "even out" lurking variables. 16. If you ran an experiment where you randomly assign people to either listen to music or silence, would you have to worry about the effect of SAT prep courses on the results? Explain.	
3. *Wrap Up* **Take-Home Messages** ➢ An experiment does not require random assignment of treatments – you just have to assign a treatment. When you randomly assign treatments, it is called a randomized experiment. ➢ This lesson is critical for understanding how experiments differ from observational studies. When we assign treatments at random, we "even out" any lurking variables, so we can say that differences we observe are caused by the explanatory variable (the treatment). We call this causal inference. Causal inference cannot be made with observational studies. ➢ Our use of the web app today was to see what happens to means of a quantitative lurking variable when we randomly split people into two groups. You should have concluded that the means tend to be approximately equal (difference in means is centered at zero), and that the distribution of the difference in means is symmetric. Any positive value has a negative counterpart, which just involves swapping the labels (T ↔ C).	~ 5 min
4. *Homework* (OPTIONAL) At your discretion, assign homework problems of your choice. For sample question, see the *obs_exp_hw_sample.docx* and the *obs_exp_hw_sample_solutions.docx* files.	



Does Music Help Us Study? – Part 1
Experiments vs. Observational Studies



Do you tend to listen to music while studying? Does it help or hinder your studying? Suppose a researcher wants to investigate the following research question:

"Do some types of music improve or hurt the effectiveness of studying?"

We will examine different studies that address this question of interest.

1. Suppose we want to address the question of interest by getting college school students to fill out a **survey**.

   a. The survey will ask for details on the music type each student prefers for studying, but we will also need a way to measure how effective their studying is. How could we measure a **response** to use for comparison – to see how much people are learning while studying?

   b. What differences in students make it hard to get a clear comparison between different music types? List at least three variables that we should consider. For each: is it categorical or quantitative?

2. Another option for studying the effect of music type on studying is to **assign treatments**, as in this study from 2014.

   Perham, N. and Currie, H (2014). Does listening to preferred music improve reading comprehension performance? *Applied Cognitive Psychology* **28**:279-284.

   They used four levels of the variable sound: "*disliked lyrical music* (DLYR), *liked lyrical music* (LLYR), *non-lyrical music* (NLYR) and *quiet* (Q)" and each subject chose music they liked with lyrics (LLYR), while the instrumental music (NLYR) was picked by the researchers, and subjects were screened to be sure they did not enjoy "thrash" music, which was used for DLYR. Subjects were told to ignore the music, and had to read 70 lines of text, then answer four multiple choice questions about the reading (taken from SAT exams). They repeated the task for three more readings (with 4 questions each), and the proportion correct was recorded.

   a. Is the use of SAT questions an improvement over your choice of response in 1a? Explain why or why not.

   b. Is the use of four sound treatments an improvement over asking students how they study? Explain why or why not.

Experiment versus Observational Study
In an **experiment**, levels of the **explanatory** (treatment) variable are **assigned** to participants (or units if people are not involved). In order to allow statistical inference, we should assign the treatments at random, making it a **randomized experiment**. In an **observational study** we simply record levels or values of the explanatory variable instead of assigning them.

3. Looking back at the studies above (problems 1 and 2), which was an experiment and which was an observational study. Explain how you know this.

Advantages of Randomized Experiments

To make sure we're all thinking of the same response for our study on the effect of music while studying, we'll focus on using the SAT reading comprehension scores as our **response**. Music (or quiet) will be played while our subjects read and answer the questions.

4. In 1b, above you mentioned several attributes of students which would indicate who does better on a test. One such variable would be IQ. People with higher IQ scores tend score higher on the SAT. We refer to a variable like IQ as a **lurking variable** when we do not measure it and take it into account. What other lurking variables did you identify in 1b (or add some here to get at least three), which would cause some students to do better on SAT than others?

5. If we don't measure IQ and don't adjust for it, we won't be able to tell whether one group did better because it had higher mean IQ, or because they were assigned the more effective treatment. Let's see what happens to mean IQ (and another variable - SAT Prep) if we randomly separate 12 students into treatment (music) and control (quiet) groups of 6 each.

Treatment (Music)			Control (Quiet – No Music)		
Name	**IQ**	**SAT Prep**	**Name**	**IQ**	**SAT Prep**
Andy	104	Yes	Peter	106	Yes
Ben	118	Yes	Maria	90	No
Betty	79	No	Marti	97	No
Jorge	94	Yes	Mikaela	98	No
Kate	106	No	Patty	89	No
Russ	88	Yes	Shawn	85	Yes

Mean IQ of Treatment Group = 98.2
Mean IQ of Control Group = 94.2
Difference in Means = 4

Write the Name, IQ score, and whether or not they took an SAT prep class (Yes or No) for each student on an index card.

a. Mix the cards thoroughly, and deal them into two piles of six each, labeling one pile "T" for Treatment and the other "C" for Control. Compute the mean IQ for each group and take the difference (Treatment – Control). *Notice that IQ score is a quantitative variable.*

- Mean IQ for Treatment Group =
- Mean IQ for Control Group =
- Difference (Treatment – Control) =

6. As with many techniques in statistics, we want to see what happens when we **repeat the process many times**. Doing this by hand many times gets tedious, so we will use the computer to shuffle and compute means for us.

Go to http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/ and click Lurking Demo under One Quant.. Select IQ (normal). It gives you a bigger sample – 25 IQ's in each group (newly generated at random from a symmetric distribution with mean 100 and standard deviation 15).

a. Write down the means for each group in the first shuffle (Randomization 1) and their difference.

- Mean for Treatment Group =
- Mean for Control Group =
- Difference (Treatment – Control) =

b. Write down the means for each group in the second shuffle (Randomization 2) and their difference.

- Mean for Treatment Group =
- Mean for Control Group =
- Difference (Treatment – Control) =

c. Compare your answers with another group's answers.

7. As we said above, we need to think about repeating the shuffling process over and over. Click 5000 "more randomizations" and sketch the right-hand plot. Describe the center (mean), spread (SE), and shape of this distribution. SE stands for standard error, which is the standard deviation of the randomization distribution.

8. Do we get the same pattern in the right-hand plot if we run another batch of shuffles, say 10,000 this time? Click IQ (normal) to reset the plot and then click 5000 "more randomizations" twice to run 10,000 shuffles. Did center (mean), shape, and/or spread (SE) change?

9. Note that some differences are not close to zero. What are the approximate largest and smallest values you obtained in 10,000 shuffles?

- largest value =
- smallest value =

10. Does randomization always make mean IQs the same between the two treatment groups? Explain.

11. Does randomization tend to balance out mean IQ in the long run, after many trials? Explain.

12. In general, how similar are group mean IQs when we randomly assign people into two groups?

13. Another lurking variable would be the fact that some people have taken a short course as an SAT prep and others haven't. *Notice that SAT prep is a categorical variable.* If the SAT prep course does what it claims, then it could be the reason for one group to score higher than the other.

    If you ran an experiment where you randomly assigned people to either listen to music or silence, would you have to worry about the effect of SAT prep courses on the results? Explain.



Does Music Help Us Study? – Part 1

Solutions

Experiments vs. Observational Studies



1. Suppose we want to address the question of interest by getting college school students to fill out a **survey**.

   a. The survey will ask for details on the music type each student prefers for studying, but we will also need a way to measure how effective their studying is. How could we measure a **response** to use for comparison – to see how much people are learning while studying?

We want them to wrestle with questions of how much subjects knew before studying, how good a student they are, etc. (more in next question). Good responses: a post-test minus pre-test difference or we pick a topic which people don't generally know about before they study.

b. What differences in students make it hard to get a clear comparison between different music types? List at least three variables that we should consider. For each: is it categorical or quantitative?

Answers will vary. Below are some possibilities.

- *Some students have higher IQ than others (IQ is quantitative)*
- *The subject area of the test they are studying for (anatomy versus philosophy?) (Subject area is categorical)*
- *Years of musical training (quantitative)*
- *Surroundings (e.g., dorm room versus library) (categorical)*
- *Age (quantitative)*

2. Another option for studying the effect of music type on studying is to **assign treatments**, as in this study from 2014.

a. Is the use of SAT questions an improvement over your choice of response in 1a? Explain why or why not.

SAT will generally be better because it allows us to use a comparable measure across all subjects, and it is immediately following the music treatment.

b. Is the use of four sound treatments an improvement over asking students how they study? Explain why or why not.

The four sound *treatment levels are generally better because they take away many of the options people have when choosing music. We can then get a direct comparison of music versus no music and of lyrical versus instrumental music.*

Experiment versus Observational Study

3. Looking back at the studies above (problems 1 and 2), which was an experiment and which was an observational study. Explain how you know this.

Problem 2 (the article by Perham and Currie) is an ***experiment*** *because four levels of the variable* sound *(i.e., music levels) were assigned to participants.*

Problem 1 (the survey) is an ***observational study*** *because music levels were NOT assigned.*

Advantages of Randomized Experiments

4. In 1b, above you mentioned several attributes of students which would indicate who does better on a test. One such variable would be IQ. People with higher IQ scores tend to score higher on the SAT. We refer to a variable like IQ as a **lurking variable** when we do not measure it and take it into account. What other lurking variables did you identify in 1b (or add some here to get at least three), which would cause some students to do better on SAT than others?

Answers will vary. Some of their responses in 1a may be lurking variables. For example: whether or not a student takes an SAT prep course, musical tastes, etc. Type of music played during the experiment is something that the researchers can control for so I would not say that this is a

possible lurking variable. The lurking variables for this example should be things about the participants that the researchers did not get to decide.

5. If we don't measure IQ and don't adjust for it, we won't be able to tell whether one group did better because it had higher mean IQ, or because they were assigned the more effective treatment. Let's see what happens to mean IQ (and another variable - SAT Prep) if we randomly separate 12 students into treatment (music) and control (quiet) groups of 6 each.

    - Mean IQ for Treatment Group = *Answers will vary*
    - Mean IQ for Control Group = *Answers will vary*
    - Difference (Treatment – Control) = *Answers will vary*

6. As with many techniques in statistics, we want to see what happens when we **repeat the process many times**. Doing this by hand many times gets tedious, so we will use the computer to shuffle and compute means for us.

    a. Write down the means for each group in the first shuffle (Randomization 1) and their difference.

    - Mean for Treatment Group = *Answers will vary*
    - Mean for Control Group = *Answers will vary*
    - Difference (Treatment – Control) = *Answers will vary*

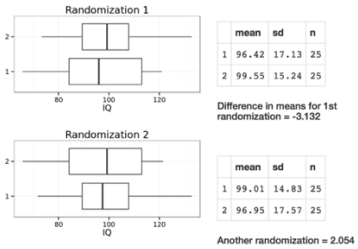
    b. Write down the means for each group in the second shuffle (Randomization 2) and their difference.

    - Mean for Treatment Group = *Answers will vary*
    - Mean for Control Group = *Answers will vary*
    - Difference (Treatment – Control) = *Answers will vary*

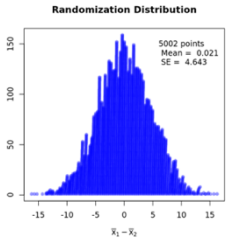
    c. Compare your answers with another group's answers.

    Answers will vary. The answers to part (a) and (b) are likely to vary between groups.

    Below is what I obtained for my first and second shuffle.

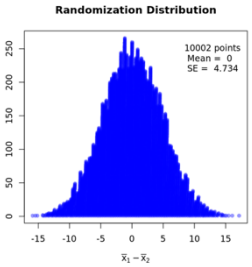


7. As we said above, we need to think about repeating the shuffling process over and over. Click 5000 and sketch the right-hand plot. Describe the center (mean), spread (SE), and shape of this distribution. SE stands for standard error, which is the standard deviation of the randomization distribution.



Center: *Near 0 (look at the mean value)*
Spread: *Around 4 (look at the SE value)*
Shape: *Symmetric*

8. Do we get the same pattern in the right hand plot if we run another batch of shuffles, say 10,000 this time? Click IQ (normal) to reset and then click 5000 "more randomizations" twice to run 10,000 shuffles. Did center (mean), shape, and/or spread (SE) change?



You should get the same pattern in the plot and the center, shape and spread all remain about the same.

9. Note that some differences are not close to zero. What are the approximate largest and smallest values you obtained in 10,000 shuffles?

   - largest value = *Answers will vary. In the plot above, the largest value is about 17*
   - smallest value = *Answers will vary. In the plot above, the smallest value is about -16.*

10. Does randomization always make mean IQs the same between the two treatment groups? Explain.

    Not every single trial, though on average (over many shuffles), the groups are equivalent.

11. Does randomization tend to balance out mean IQ in the long run, after many trials? Explain.

    Yes, because the distribution of the difference between the two group means is centered at 0.

12. In general, how similar are group mean IQs when we randomly assign people into two groups?

    The difference in mean IQ between two randomized groups will be close to zero most of the time. There is no guarantee for any one sample, but in general randomization makes the groups very similar.

13. Another lurking variable would be the fact that some people have taken a short course as an SAT prep and others haven't. *Notice that SAT prep is a categorical variable.* If the SAT prep course does what it claims, then it could be the reason for one group to score higher than the other.

    If you ran an experiment where you randomly assigned people to either listen to music or silence, would you have to worry about the effect of SAT prep courses on the results? Explain.

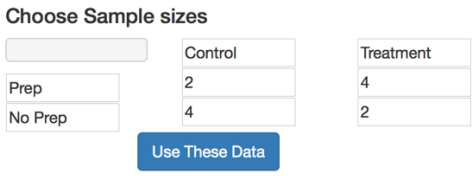
We can be pretty sure that the two randomized groups have similar proportions of people in them who took the SAT prep course. Therefore, it's not a problem, and we can make out conclusions on the effects of music without worry about lurking variables.

Additional Problems - Does Music Help Us Study? – Part 1 Activity

14. Another lurking variable would be the fact that some people have taken a short course as an SAT prep course and others haven't. If the SAT prep course does what it claims, then it could be the reason for one group to score higher than the other. We will look at the proportions who have taken an SAT prep course in the treatment and control groups. We are now looking at proportions instead of means because this lurking variable (took SAT Prep course) is a categorical variable.

    a. Compute the proportion of Yes's in the two groups of cards you had shuffled, and subtract. Write the proportions and the difference here.

    b. Again go to http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/ and click Lurking Demo, but this time under the One Categ. header. Enter the data as follows:

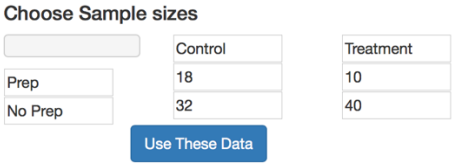


    and click Use These Data. Run 5000 "more randomizations" and record the mean and SE of the differences.

    - Mean =
    - SE =

    c. You will have a few shuffles that give -1 or 1. How could that happen?

    d. The plot gets more interesting with larger counts. Suppose we are randomly assigning 100 people to our two groups, and that 28 of them have taken an SAT prep course, 72 have not. Of those in the treatment group (Music), 10 of them have taken an SAT Prep course and 40 have not. Of those in the control group (No Music), 18 have taken an SAT Prep course and 32 have not. Enter the data as follows:



    and click Use These Data. What proportions of "Prep" are in the treatment group and what differences do you get for the first two randomizations?

    e. Run 5000 "more randomizations." Sketch your plot here.

f. Compare with other groups. Do the pictures look the same?

15. When we randomly assign people to two groups:

    a. Is it possible for a categorical lurking variable like SAT prep to be imbalanced across the two groups? Explain.

    b. Will the lurking SAT prep variable "usually" be poorly balanced across the two groups? Explain?

16. In general, how similar is the proportion of people who have taken SAT prep in the treatment group to the same proportion in the control group?

17. If you ran an experiment where you randomly assign people to either listen to music or silence, would you have to worry about the effect of SAT prep courses on the results? Explain.

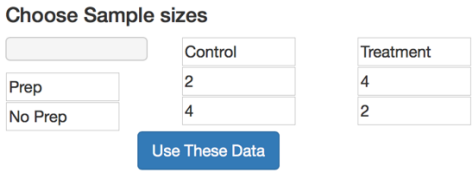
Additional Problems - Does Music Help Us Study? – Part 1 Activity
SOLUTIONS

14. Another lurking variable would be the fact that some people have taken a short course as an SAT prep course and others haven't. If the SAT prep course does what it claims, then it could be the reason for one group to score higher than the other. We will look at the proportions who have taken an SAT prep course in the treatment and control groups. We are now looking at proportions instead of means because this lurking variable (took SAT Prep course) is a categorical variable.

    a. Compute the proportion of Yes's in the two groups of cards you had shuffled, and subtract. Write the proportions and the difference here.

    Answers will vary. Make sure they record the proportions of Yes's in each group and then subtract (Group 1 – Group 2).

    b. Again go to http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/ and click Lurking Demo, but this time under the One Categ. header. Enter the data as follows:



    and click Use These Data. Run 5000 "more randomizations" and record the mean and SE of the differences.

    - Mean = *Answers will vary, but the mean should be close to 0.*
    - SE = *Answers will vary, but the SE should be about 0.3.*

    c. You will have a few shuffles that give -1 or 1. How could that happen?

One group got all the Prep people and the other group got none.

d. What proportions of "Prep" are in the treatment group and what differences do you get for the first two randomizations?

Answers will vary. Below is what I obtained.

First Randomization

	Prep	n	Prop
Control	18	50	0.360
Treatment	10	50	0.200

Difference in proportions: 0.16

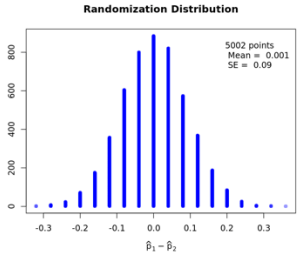
Another Randomization

	Prep	n	Prop
Control	14	50	0.280
Treatment	14	50	0.280

Difference in proportions: 0

e. Run 5000 "more randomizations." Sketch your plot here.

Answers will vary. Below is an example of what the plot should look like.



f. Compare with other groups. Do the pictures look the same?

The pictures should look the same. The mean will be near 0. The SE (spread) will be approximately 0.09 and the shape of the randomization distribution should be symmetric.

15. When we randomly assign people to two groups:

a. Is it possible for a categorical lurking variable like SAT prep to be imbalanced across the two groups? Explain.

Yes, the proportions might differ by as much as 0.3.

b. Will the lurking SAT prep variable "usually" be poorly balanced across the two groups? Explain?

No. For most of the randomizations, we get a difference in proportion which is close to zero, which says that about the same proportion of treatment people as control people have taken an SAT pre course.

16. In general, how similar is the proportion of people who have taken SAT prep in the treatment group to the same proportion in the control group?

    The difference in proportions with SAT prep between two randomized groups will be close to zero most of the time. There is no guarantee for any one sample, but in general randomization makes the groups very similar.

17. If you ran an experiment where you randomly assign people to either listen to music or silence, would you have to worry about the effect of SAT prep courses on the results? Explain.

    We can be pretty sure that the two randomized groups have similar proportions of people in them who took the SAT prep. Therefore, it's not a problem, and we can make out conclusions on the effects of music without worry about lurking variables.

Experiments vs. Observational Studies Homework

Sample Problems

1. For his statistics class experiment, researcher J. Gilbert decides to study how parents' income affects children's performance on standardized tests like the SAT. He proposed to collect information from a random sample of test takers and examine the relationship between parental income and SAT score. What type of study was conducted (an observational study or experiment)? Explain.

2. Eighteen mice were randomly divided into two cages, one of which had a normal light-dark schedule and the other, which experienced light 24 hours per day. The weight gain, in grams, was recorded for each mouse.

    a. What type of study was conducted (an observational study or experiment)? Explain.

    b. Some mice are naturally more active than others, which would lead them to exercise more. For comparing the two groups, activity level is
        i. the explanatory variable.
        ii. the response variable.
        iii. a lurking variable that will cause bias in the results.
        iv. a lurking variable that should get evened out by randomization.

3. 50 randomly selected athletes who had suffered sprained ankles were randomly assigned to one of two exercise programs. The first group returned to sports activities in a mean of 38.7 days. The other group returned to sports in a mean of 18.3 days. Why is it important to assign the athletes to the two treatments randomly?

4. What is the primary difference between observational studies and experiments?
    a. Observational studies will not have lurking variables.
    b. Experiments must use a random sample.
    c. Experiments require assignment of participants to treatments.

Experiments vs. Observational Studies Homework

SOLUTIONS

Sample Problems

1. For his statistics class experiment, researcher J. Gilbert decide to study how parents' income affects children's performance on standardized tests like the SAT. He proposed to collect information from a random sample of test takers and examine the relationship between parental income and SAT score. What type of study was conducted (an observational study or experiment)? Explain.

   Observational study because there were no treatments enforced. The researcher simply observed and recorded the parental income and SAT score as they naturally exist.

2. Eighteen mice were randomly divided into two cages, one of which had a normal light-dark schedule and the other, which experienced light 24 hours per day. The weight gain, in grams, was recorded for each mouse.

   a. What type of study was conducted (an observational study or experiment)? Explain.

      Experiment because the rats were assigned a treatment (normal light-dark schedule or light 24 hours per day).

   b. Some mice are naturally more active than others, which would lead them to exercise more. For comparing the two groups, activity level is
      i. the explanatory variable.
      ii. the response variable
      iii. a lurking variable that will cause bias in the results.
      iv. a lurking variable that should get evened out by randomization.

      Correct answer: iv

3. 50 randomly selected athletes who had suffered sprained ankles were randomly assigned to one of two exercise programs. The first group returned to sports activities in a mean of 38.7 days. The other group returned to sports in a mean of 18.3 days. Why is it important to assign the athletes to the two treatments randomly?

   In order to even out the lurking variables between the two groups so that causal inferences (i.e., any differences observed are caused by the treatment) can be drawn.

4. What is the primary difference between observational studies and experiments?
   a. Observational studies will not have lurking variables.
   b. Experiments must use a random sample.
   c. Experiments require assignment of participants to treatments.

   Correct answer: c

Overview of Lesson 4

<table>
<tr><td colspan="2">Topic of Lesson:
Interval Estimation and Margin of Error</td><td>Date:</td></tr>
<tr><td colspan="3">Summary of Lesson:
This lesson centers on interval estimation and margin of error. The activity uses data from a sample survey and students are asked to use information to estimate a population proportion. Next, the activity guides students through the idea of variability within sample proportions and how an interval estimate (opposed to a point estimate) can describe more of a “story.” A web applet is used to conduct a simulation to investigate how variable a sample proportion might be. Students are introduced to the concept of bootstrapping and run a simulation to develop a margin of error and interval estimates. Finally, 95% confidence intervals are constructed using the web app (“percentile method”) and by hand (“Estimate $\pm$ 2 $\times$ SE method”).

Note: This web applet works on computers, tablets, and smartphones.</td></tr>
<tr><td>Prerequisites:
• None</td><td>Activities:
Does Music Help Us Study? – Part 2</td><td>Estimated Time:
Approximately 100 (depends on the amount of time you want to spend on the activity)</td></tr>
<tr><td colspan="2">Learning Objectives
Students will be able to …
• understand what an interval estimate and margin of error is.
• develop an interval estimate and margin of error through the use of simulation.
• recognize the advantage of using an interval estimate over a point estimate.
• learn what bootstrapping is and why and how we use this method.</td><td>Assessments
• Optional HW assignment with problems of your choice (see int_estimate_ME_hw_sample.docx file for possible homework problems)</td></tr>
<tr><td colspan="2">CCSSM Standards Addressed
• S.IC.4. Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.</td><td>CCSSM Mathematical Practices Emphasized
• MP1. Make sense of problems and persevere in solving them.
• MP3. Construct viable arguments and critique the reasoning of others.
• MP5. Use appropriate tools strategically.</td></tr>
<tr><td colspan="3">Teaching Materials
• Copies of the Does Music Help Us Study? – Part 2 activity (one per student or one per group)</td></tr>
</table>

- Access to a computer, tablet, or smartphone
- Completed sets (one per group or one per student) of cards for #8 on the activity OR blank sets of cards that students can write information on themselves.

Learning Plan for Lesson 4

Teacher Activity	Estimated Time
1. *Introduction to Margin of Errors* (OPTIONAL) An optional introduction about an interval estimate and margin of error can be implemented prior to the activity. For instance, to motivate this topic, you can provide students with a real-life example of a margin of error and/or interval estimate reported in a study and ask a question about it. The most typical examples reporting a margin of error and/or an interval estimate include poll results. *Example:* *On January 30-31, 2007, Fox News/Opinion Dynamics polled 900 registered voters nationwide. When asked, "Do you believe global warming exists?" 82% said "Yes." Fox reported their margin of error to be ±3%.* ***Question:*** *It is standard among pollsters to use a 95% confidence level unless otherwise stated. Given that, what does Fox News mean by claiming a margin of error of ±3% in this context?* ***Answer:*** *(Answer can be discussed at any time (e.g., now or after the activity is completed)* *If this polling was done repeatedly, 95% of all random samples would yield estimates that come within ±3% of the true proportion of all registered voters who believe that global warming exists.*	~ 5 min
2. *Does Music Help Us Study – Part 2 Activity*	

Pass out the activity. Access to a computer, tablet, or smart phone will be needed to use the web applet mentioned in the activity. Note that the use of this web applet does not occur until Problem #9 in the activity. Problems 1 – 8 do not require the use of the applet. It's recommended to let students work through this activity together in groups of 2 – 3 where each student or each group uses a computer/tablet/smartphone. Groups of 2 using 1 computer (and working together) have worked out well in the past. If you have each student use a computer/tablet/smartphone, note that they will get slightly different answers than others when using the web applet. This may frustrate those working together, but the web applet generates data anytime it is used which results in some variability. The activity also works best when students discuss and think about their answers together as opposed to solely completing the activity without any discussion. This activity uses the same web applet from the previous lesson and students tend to become more comfortable with this app each time they use it. **Problems 1 – 3** ➢ These questions set up the design of the study (i.e., the response variable and its type (categorical or quantitative) and the parameter (population proportion) that the researcher is interested in). Letting students work through these questions as a group and then having a large group discussion or individually checking in with each group works best. Of these questions, #3 may be the most difficult for them to answer. Students may focus on the sample of students rather than the entire population of rural high school students in the U.S. The key is that this answer should be about the population. Additional notation and terminology can be described at this point if you prefer. For example, the population proportion that is being described can be called a parameter and the notation used for this parameter is p. **Problems 4 – 5** ➢ These questions guide students to summarize the data provided at the beginning of the activity (i.e., identify the sample size and statistic (sample proportion)).	~ 90 min (Estimated time depends on how the activity is competed. For example, working through the problems and applet without any difficulties or discussions can take about 35 minutes. Discussing problems and troubleshooting any difficulties adds additional time.)

For question #4, some students might respond that the sample size is 23. Ask them to focus on the number of students who were asked the question, rather than the number of students who answered the question one way or the other.

Additional notation and terminology can be described in #5 – a sample proportion (i.e., statistic) is being calculated and the notation used is $\hat{p}$ (p-hat).

Problems 6 – 7

- These questions guide students to an important point about samples. Not all samples are the same, which tells us that we should expect some variation in our sample proportion.

These two questions are very important! Variability is central to statistics It's recommended to make sure that all students understand these answers before moving on with the remainder of the activity.

In #6, some students might not realize that a different set of 30 high school students are being asked the question.

From past experience, many students tend to skim over the notes and descriptions within the activities (e.g., the 3 paragraphs after #7). Emphasize the importance of reading and highlighting main concepts in these paragraphs throughout the activities because they provide valuable information.

Problem 8

- #8 begins the "simulation" part of this activity, but by hand. Students need 30 cards (one for each participant in the study) that have the outcomes written on it. In other words, 23 cards should be marked with an "M" (study with music playing) and 7 cards should be marked with an "S" (study in silence). *To save time during this activity, it is recommended to have these cards completed and ready beforehand. However, you may also choose to have students complete these cards themselves. These cards can be created using any time of paper.*

Graphical addition to Problem #8 & a great question to ask regarding the plot

Similar to #5a from the previous activity (*Does Music Help Us Study – Part 1*), you can ask students to report their proportion and create a re-sampling distribution plot using the class data. It is suggested to create a dot plot of the proportions and discuss the plot as a group. This gives a nice introduction to what will be seen using the web applet (#9 & 10), but on a much smaller scale. For example, GeoGebra (www.geogebra.org) can be quickly used to make a dot plot. Once you at at the website, you would click on "Start GeoGebra", select "Spreadsheet", enter the differences in column A, select column A, click on the purple histogram at the top of the window and select "One Variable Analysis." A histogram will automatically appear and if you click on

the double arrows after Histogram at the top of the plot, you will see the option for a "Dot Plot." Additional instructions can be found using a quick Google search.

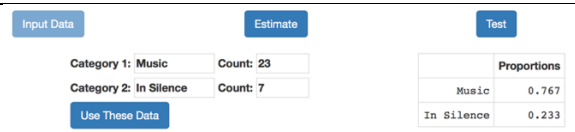
By looking at just the class data, students can better understand how the re-sampling plot is created. A question we often ask students about this type of plot is **"How is one dot created?"** ***This is a very hard question for students and they often struggle with it (even if you have previously asked this question)."*** For example, students may refer back to how they previously answered this question and say "It's the difference in means." or "It's a data point." Having students write down what they had to do by hand in #8 sometimes helps students with this "one dot" question. The answer, in the context of what students did by hand and what they will do with the web applet, is "30 cards were created and marked with "M" or "S" to represent the outcomes from the original sample taken. That is, 23 cards were marked with "M" and 7 were marked with "S". We then sampled with replacement 30 times from this original sample and plotted the proportion of participants who prefer to study with music playing."

Furthermore, asking this type of question helps set up students to make connections later on between what they did by hand and what the web applet is doing. Both are doing the same thing, but the web applet is much faster.

Problems 9 – 12

- Access to a computer, tablet, or smartphone is needed.
- The web applet used for this activity (http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/) is the same app used in Lesson 3. The directions in #9 help set up the app/demonstration and asks what proportion of students who prefer to study with music is simulated. The concept of bootstrapping is introduced right before this problem.
- #10 – 12 ask questions regarding the use of the web applet and range from creating a re-sampling distribution, investigating different confidence levels, and constructing an interval estimate and margin of error.

#9 starts the use of the web applet. Students should be able to follow the directions to set up the data. From previous experience, many students have forgotten to click Use These Data. Make sure students are clicking this before moving on (otherwise they won't be able to use the Estimate feature on the web applet). You can tell that the data has been entered correctly and that Use These Data has been clicked because a table with calculated proportions (as shown below) will appear on the app. If this does not appear, then Use These Data has NOT been clicked.

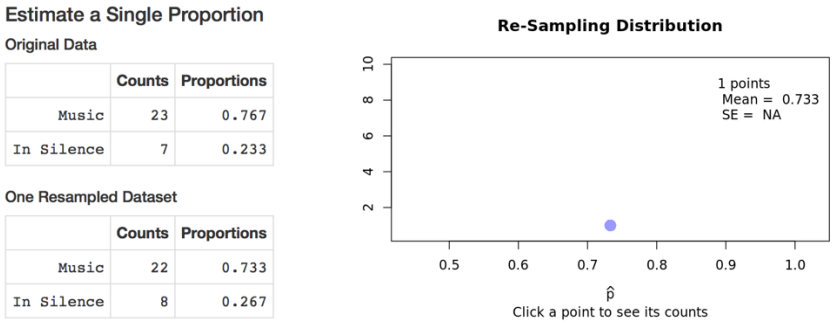


In addition, if students fail to click on Use These Data, then they will not even be able to click on Estimate. You'll notice that Estimate stays a slightly lighter shade of blue until Use These Data has been selected.

********** IMPORTANT **********

The question in #9 (after they set up the dataset) asks students to report the proportion that prefer to study with music - students may have a hard time seeing what this value is because a summary table does not automatically show up. There are two ways to find this value: (1) Since there is initially only 1 dot in the re-sampling distribution, the value of this dot can be found by looking at the mean stated in the top right of the plot. This is the proportion that they are looking for. (2) A very cool attribute of this plot in the web applet is to CLICK ON the single dot in the plot. You can use this feature on any dot once more have been simulated. When you click on a dot, a table will appear below "One Resampled Dataset" (see picture below) and the proportions are displayed for the Music and In Silence groups. Being able to click on a particular dot is a great attribute to use to analyze individual dots when you have hundreds or thousands of values plotted!

****



Note that the counts in both summary tables sum to 30! If this is not the case, the data has been entered incorrectly.

The value of the proportion in the Music group will likely differ for everyone. It's possible to get a proportion of 1 – if this occurs, ask students what that means in the context of the problem. This is another way to ask "How is one dot created?" If a student received a proportion of 1, then it means that all 30 of the participants preferred to study with music playing. If the proportion was -1, then all 30 of the students preferred to study in silence.

#10 and #11 asks the students to generate 100 additional resamples. For #10, it's okay if they accidentally generated more than 100 resamples. Their answers will be approximately the same. However, if they generated more than 100 resamples for #11, they could still complete the question, but may have a difficult time physically counting the dots. In any case, to reset, a student needs to click on Input These Data near the top of the screen. The data should still be entered and students can click on Use These Data and then click Estimate to return to the same page as before. Students should see that the plot has been reset.

#10c is an important question to address and is also a question that students struggle with. They typically are not sure why it's centered at this value.

#11 – Parts a. through c. are straightforward. When answering part d. it's helpful to remind students that the total number of dots is 101. Some students use 100 as the total number of dots which gives slightly different percentages.

Addition to #11 – Web Applet

After students make their prediction in #11f, you can ask them to verify (using the app) whether or not their prediction is correct. Have students generate 100 more re-samples by clicking 100 "more resamples" again and selecting a 90 confidence level %. Now, they can compare the number of blue and red dots with their predictions.

#12 – students should be able to follow the directions to construct a 95% confidence interval by hand and with the web app. Emphasize that margin of error is what you add or subtract to your estimate. In other words, the margin of error is the half-width of the confidence interval created. Narrower intervals are more ideal! When calculating a confidence interval, a helpful tip to provide students is to always subtract the margin of error first – this will give the lower endpoint of the interval – and then add the margin of error – this will give the upper endpoint of the interval. Students have switched up the endpoints.

Addition to #12 – Interpreting confidence intervals

Since #12 has students create a confidence interval, you can further address confidence intervals by explaining how one can interpret them. For instance, *"our confidence in a particular 95% confidence interval is actually in the process used to create the interval. We know that using this process over and over again gives intervals, which will usually (95% of the time) cover the true value.* ***Therefore, we can say that we are 95% confident that the true proportion of rural high school students in the U.S. who prefers to study with music playing is between # and #.***"

3. *Wrap Up* **Take-Home Messages** ➢ We only get one sample, but we can create many "resamples" using sampling with replacement (also called bootstrapping). ➢ Interval estimates are better than point estimates. ▪ They don't pretend to be exact. Any exact value is almost certainly wrong. They describe more of the story and are able to address research questions asking "how large … , how small … , etc." where a range of values is preferred. ▪ By looking at the width of an interval (twice the margin of error) we can evaluate the quality of the data. Wide intervals are not very useful. Skinny intervals are more informative. Therefore, a large margin of error is less useful than a smaller margin of error. ▪ Our methods are not "fail safe", but are actually designed to have a certain error rate.	~ 5 min
4. *Homework* (OPTIONAL) At your discretion, assign homework problems of your choice. For sample question, see the *int_estimate_ME_hw_sample.docx* and the *int_estimate_ME_hw_sample_solutions.docx* files. Some of these questions involve the use of the web applet.	



Does Music Help Us Study? – Part 2
Interval Estimate for a Proportion



A researcher is investigating whether students in rural high schools in the U.S. prefer to study with music playing or study in silence. The researcher takes a sample of 30 randomly selected students and asks them the following question:

"Do you prefer to study with music playing or study in silence?"

Of the 30 students in the study, 23 of them stated that they prefer to study with music playing. Because the students in this study were randomly selected, they can be assumed to be a "representative sample" from the population of rural high school students in the U.S.. The researcher needs a good estimate of the true proportion of these students who prefer to study with music playing.

Step 1: State the research question.

The researcher's main question of interest is

"How large is the proportion of rural high school students who study with music playing?"

Step 2: Design a study and collect data.

1. What responses will the researcher record from each student in the study?

2. Your answer above gives the outcome of the variable of interest in this study. Is this variable quantitative or categorical?

3. What is the population proportion the researcher is interested in? Describe it in words.

Step 3: Explore and summarize the data.

4. What is the sample size in this study?

5. Calculate the proportion of students in the sample who prefer to study with music playing.

6. If the study was repeated with another 30 high school students, do you think you would get exactly 23 again who prefer to study with music playing? Explain.

Step 4: Draw inferences beyond the data.

7. The point in the last question is really important. When we repeat the same study, we do not get exactly the same results. Why is that?

We know exactly what proportion of rural high school students in the sample prefer to study with music playing, and that number makes a good estimate of the same proportion of all rural high school students who prefer to study with music playing. However, the fact that not all students, and not all samples are the same tells us we need to expect some variation in our sample proportion when we repeat the study.

A single number like the one you computed in #5 does not tell the whole story. We want to know "how good" this estimate is. One way to report the quality of an estimate is to give a range of values – an interval estimate – instead of a single "point estimate.

Because we now have easy access to computers, we can run a **simulation** to see how variable the sample proportion might be. We only get one sample of real data, but we can create lots of simulated datasets, which represent other values, which might have been observed.

8. Your group will get 30 cards on which you will write the observed outcomes of the study. That is, write "M" for *study with music playing* on 23 cards and write "S" for *study in silence* on 7 cards. Next, we simulate the study on another sample of 30 students. However, we don't have the time to ask more students about their music preference while studying, so we "recycle" the numbers we have.

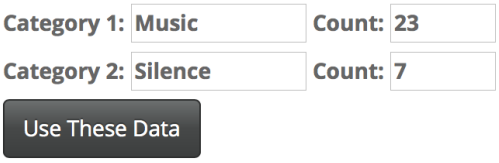
    a. Shuffle your cards and draw one at random. Record the outcome below (i.e., write down "M" or "S"). Replace the card into the deck. This is a simple but powerful idea. By sampling **with replacement** we have the same conditions for every draw, and the probability of each outcome stays the same. Shuffle, draw a card, and record the outcome. Repeat until you have 30 outcomes recorded and chosen at random.

    b. What proportion of your students prefers to study with music playing?

The process you just used is called **bootstrapping** (which means to make something out of little or nothing), and the 30 outcomes are called a bootstrap **resample**. It's not a sample – we only get one of those – and we can repeat the resampling process many times.

9. Reshuffling by hand is slow, so we want to speed up the process by using technology. Our goal is to see what other outcomes we might have gotten for different samples of 30 students.

    We will use the One Categ. web app at http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/. Select Test or Estimate under One Categ. to enter the music data to look like:



    Then click on Use These Data. Now, click Estimate. A single dot should be on your graph. Click on that dot and a table will appear on the left side of the window under "One Resampled Dataset." Looking at this table, what proportion of the students prefers to study with music?

10. Now resample 100 times by clicking 100 "more resamples."

    a. Where is the re-sampling distribution centered (i.e., what is the mean?)

    b. How spread out are the sample outcomes? (i.e., what is the SE? SE stands for standard error, which is the standard deviation of the resampled values.)

    c. The center should seem reasonable. Why is the distribution centered at this value?

11. You should have 101 resampled values in your plot. Below the plot are options for confidence levels for our interval estimate.

a. Click 80 "confidence level %" and count:
   i. How many red points in the left tail?
   ii. How many red points in the right tail?
   iii. How many blue points in the middle?

b. Click 90 "confidence level %" and count:
   i. How many red points in the left tail?
   ii. How many red points in the right tail?
   iii. How many blue points in the middle?

c. Click 95 "confidence level %" and count:
   i. How many red points in the left tail?
   ii. How many red points in the right tail?
   iii. How many blue points in the middle?

d. Explain how the confidence level % is related to the number of blue points.

e. Play with the Confidence Level % buttons more to explain: How are the endpoints of the interval estimate related to the colors of the points in the plot?

f. Predict: What will happen to the numbers in each tail for, say, a 90% interval, if we double the number of resamples?

In problem #11, you clicked on a confidence level and the web app colored in the right number of dots as red to put our selected percentage of sampled proportions in the center (these stayed blue) and split the remainder into the two tails, turning these more extreme points red. We call this a "percentile" method because, for example, a 90% confidence interval has a lower endpoint of the 5th percentile (5% of the values below) and an upper endpoint of the 95th percentile (95% of the values below).

Another common way of building a 95% confidence interval is to take the estimated value and add and subtract twice the standard error of the statistic. A 95% CI is then

$$Estimate \pm 2 \times SE$$

where the SE is a number coming from the plot on the web app. Why 2? Well, it's easy to remember, and with a symmetric bell-shaped distribution, 95% of the data will fall within 2 standard deviations of the mean. **Margin of error is then the amount we add and subtract.** In this case, it is twice the SE.

12. Go back to #9 and re-enter the data as you did in this problem. Click Input Data near the top of the window. The data should already be entered as before and make sure to click Use These Data again. Now click on Estimate and generate 5,000 to 10,000 more resamples by clicking on 5000 "more resamples" one or two more times.

    a. Select a 95 "confidence level %" and write down the interval.

b. Write down the SE shown near the top right corner of the plot. (We will not use the mean of the plotted values).

c. Add and subtract the margin of error ($2\times SE$) from the original proportion of students who preferred to study with music playing (i.e., 23/30 = 0.767) and write your answer in interval notation (e.g., (#, #)). Show your work.

d. Compare your interval in part a. to your interval in part c. Is one wider? Is there a shift?



Does Music Help Us Study? – Part 2
Solutions
Interval Estimate for a Proportion



Step 2: Design a study and collect data.

1. What responses will the researcher record from each student in the study?

   Study with music or study in silence.

2. Your answer above gives the outcome of the variable of interest in this study. Is this variable quantitative or categorical?

   Categorical.

3. What is the population proportion the researcher is interested in? Describe it in words.

   The true proportion of rural high school students who prefer to study with music playing.

Step 3: Explore and summarize the data.

4. What is the sample size in this study?

   30

5. Calculate the proportion of students in the sample who prefer to study with music playing.

   $\frac{23}{30} = 0.767$

6. If the study was repeated with another 30 high school students, do you think you would get exactly 23 again who prefer to study with music playing? Explain.

   Answers can vary.
   Not exactly 23 - a few more or less than 23 would prefer to study with music playing.

Step 4: Draw inferences beyond the data.

7. The point in the last question is really important. When we repeat the same study, we do not get exactly the same results. Why is that?

   High school students differ from each other. We've seen that not all 30 high school students in our sample have the same preference, so if we randomly grab another 30 high school students, we expect either fewer or more to prefer studying with music playing, not always 23.

8. Your group will get 30 cards on which you will write the observed outcomes of the study. That is, write "M" for *study with music playing* on 23 cards and write "S" for *study in silence* on 7 cards. Next, we simulate the study on another sample of 30 students. However, we don't have the time to ask more students about their music preference while studying, so we "recycle" the numbers we have.

   a. Shuffle your cards and draw one at random. Record the outcome below (i.e., write down "M" or "S"). Replace the card into the deck. This is a simple but powerful idea. By sampling **with replacement** we have the same conditions for every draw, and the probability of each outcome stays the same. Shuffle, draw a card, and record the outcome. Repeat until you have 30 outcomes recorded and chosen at random.

      Answers will vary.

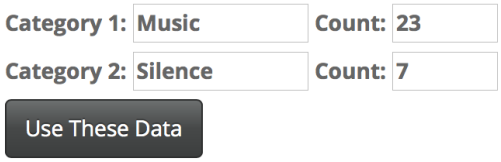
   b. What proportion of your students prefers to study with music playing?

      Answers will vary. Make sure they are writing a proportion down.

The process you just used is called **bootstrapping** (which means to make something out of little or nothing), and the 30 outcomes are called a bootstrap **resample**. It's not a sample – we only get one of those – and we can repeat the resampling process many times.

9. Reshuffling by hand is slow, so we want to speed up the process by using technology. Our goal is to see what other outcomes we might have gotten for different samples of 30 students.

   We will use the One Categ. web app at http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/. Select Test or Estimate under One Categ. to enter the music data to look like:

   Then click on Use These Data. Now, click Estimate. A single dot should be on your graph. Click on that dot and a table will appear on the left side of the window under "One Resampled Dataset." Looking at this table, what proportion of the students prefers to study with music?

   Answers will vary. I obtained 21/30 = 0.700.

One Resampled Dataset

	Counts	Proportions
Music	21	0.700
Silence	9	0.300

10. Now resample 100 times by clicking 100 "more resamples."

    a. Where is the Re-Sampling distribution centered (i.e., what is the mean?)

    Answers will vary. Approximately 0.767 (our value from #5)

    b. How spread out are the sample outcomes? (i.e., what is the SE? SE stands for standard error, which is the standard deviation of the resampled values.)

    Answers will vary. Approximately 0.078.

    c. The center should seem reasonable. Why is the distribution centered at this value?

    The simulation is run assuming high school students prefer to study with music playing with probability 0.767.

11. You should have 101 resampled values in your plot. Below the plot are options for confidence levels for our interval estimate.

    a. Click 80 "confidence level %" and count:
        i. How many red points in the left tail? *10*
        ii. How many red points in the right tail? *10*
        iii. How many blue points in the middle? *81*

    b. Click 90 "confidence level %" and count:
        i. How many red points in the left tail? *5*
        ii. How many red points in the right tail? *5*
        iii. How many blue points in the middle? *91*

    c. Click 95 "confidence level %" and count:
        i. How many red points in the left tail? *2*
        ii. How many red points in the right tail? *2*
        iii. How many blue points in the middle? *97*

    d. Explain how the confidence level % is related to the number of blue points.

    With 101 points, the percentage coverage approximately matches the number of points left as blue in the middle. For example, 81/101 ≈ 80%, 91/101 ≈ 90%, and 97/101 ≈ 95%.

e. Play with the Confidence Level % buttons more to explain: How are the endpoints of the interval estimate related to the colors of the points in the plot?

The endpoints are the values at which colors change from red to blue or vice versa.

f. Predict: What will happen to the numbers in each tail for, say, a 90% interval, if we double the number of resamples?

Counts should double, but the percentages stay the same.

12. Go back to #9 and re-enter the data as you did in this problem. Click Input Data near the top of the window. The data should already be entered as before and make sure to click Use These Data again. Now click on Estimate and generate 5,000 to 10,000 more resamples by clicking on 5000 "more resamples" one or two more times.

a. Select a 95 "confidence level %" and write down the interval.

(0.60, 0.90)

b. Write down the SE shown near the top right corner of the plot. (We will not use the mean of the plotted values).

0.077

c. Add and subtract the margin of error ($2\times SE$) from the original proportion of students who preferred to study with music playing (i.e., 23/30 = 0.767) and write your answer in interval notation (e.g., (#, #)). Show your work.

$0.767 \pm 2\times 0.077 = 0.767 \pm 0.154 = (0.613, 0.921)$

d. Compare your interval in part a. to your interval in part c. Is one wider? Is there a shift?

The percentile confidence interval (part a) is shifted slightly to the left and is slightly wider.

Interval Estimation and Margin of Error Homework

Sample Problems

1. In an August 2012 Gallup survey of 1012 randomly selected U.S. adults, 53% said that they were dissatisfied with the quality of education students receive in kindergarten through grade 12. They also report that the "margin of error is plus or minus 4%" for a 95% confidence interval. What is the standard error of the sample proportion?

A. 2%
B. 1%
C. 8%
D. 4%

2. A random sample of 169 Montana births was obtained after removing multiple births. The researchers found that 35 of the babies were delivered by Caesarean section (C-section). They want to estimate the true proportion of births via C-section.

    a. What is the sample size in this study?

    b. Calculate the proportion of births via C-section in the sample.

    c. How many of the babies in the sample were NOT delivered via C-section?

    d. Use the web applet at http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/. Select Test or Estimate under One Categ. Now, enter the data as follows:



    and select Use These Data. Now, click on Estimate and a single dot should be on your graph. Click on that dot and a table will appear on the left side of the window under "One Resampled Dataset." Looking at this table, what proportion of the births was delivered via C-section?

    e. Now resample 10,000 times by clicking 5000 "more resamples" twice.

        i. Where is your re-sampling distribution centered?

        ii. Why is the distribution centered at this value?

        iii. What is the standard error of the sample proportions?

        iv. Build a 95% confidence interval by hand. In other words, add and subtract $2 \times SE$ from the original proportion of births delivered via C-section and write your answer in interval notation (e.g., (#, #)). Show your work.

3. Suppose the researcher miscalculated the sample size in Problem #2. If we changed the sample size to be 338 Montana births and found that 70 of the babies were delivered via C-section, then identify all changes and explain why they change. Try it in the web applet!

    a. Would the proportion of births via C-section in the sample change?

    b. If we again created a re-sampling distribution with 10,000 resamples, would the spread in the re-sampling distribution plot from #2.e.iii change?

    c. Would the margin of error of our 95% confidence interval change?

Interval Estimation and Margin of Error Homework
Solutions
Sample Problems

1. In an August 2012 Gallup survey of 1012 randomly selected U.S. adults, 53% said that they were dissatisfied with the quality of education students receive in kindergarten through grade 12. They also report that the "margin of error is plus or minus 4%" for a 95% confidence interval. What is the standard error of the sample proportion?

   E. 2%
   F. 1%
   G. 8%
   H. 4%

   Correct answer: A → $ME = 2 \times SE \Rightarrow 4 = 2 \times SE \Rightarrow SE = 2\%$

2. A random sample of 169 Montana births was obtained after removing multiple births. The researchers found that 35 of the babies were delivered by Caesarean section (C-section). They want to estimate the true proportion of births via C-section.

   a. What is the sample size in this study?

      169

   b. Calculate the proportion of births via C-section in the sample.

      $\frac{35}{169} = 0.207$

   c. How many of the babies in the sample were NOT delivered via C-section?

      169 – 35 = 134

   d. What proportion of the births was delivered via C-section?

      Answers will vary. I obtained 35/169 = 0.207.

   One Resampled Dataset

	Counts	Proportions
C-Section	35	0.207
No C-Section	134	0.793

   e. Now resample 10,000 times by clicking 5000 "more resamples" twice.

      A. Where is your re-sampling distribution centered?

         Answers will vary. Should be centered about 0.207.

      B. Why is the distribution centered at this value?

         The simulation is run assuming babies are delivered via C-section with probability 0.207.

C. What is the standard error of the sample proportions?

Answers will vary. Should be about 0.031.

D. Build a 95% confidence interval by hand. In other words, add and subtract $2\times SE$ from the original proportion of births delivered via C-section and write your answer in interval notation (e.g., (#, #)). Show your work.

Answers will vary. Make sure they are doing this by hand rather than selecting the 95% confidence level on the web applet. They should be using 0.207 as their estimate and then adding and subtracting $2\times SE$.

$0.207 \pm 2\times 0.031 = 0.207 \pm 0.062 = (0.145, 0.269)$

3. Suppose the researcher miscalculated the sample size in Problem #2. If we changed the sample size to be 338 Montana births and found that 70 of the babies were delivered via C-section, then identify all changes and explain why they change. Try it in the web applet!

The data should be entered as such:



a. Would the proportion of births via C-section in the sample change?

No, it would still be 0.207 (70/338 = 0.207).

b. If we again created a re-sampling distribution with 10,000 resamples, would the spread in the re-sampling distribution plot from #2.e.iii change?

Yes, it gets smaller. It's now about 0.022.

c. Would the margin of error of our 95% confidence interval change?

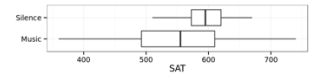
Yes, it will be smaller which means the width of the 95% confidence interval would become narrower. ME = $2 \times 0.022 = 0.044$ *instead of ME =* $2 \times 0.031 = 0.062$ *as in #2.*

Overview of Lesson 5

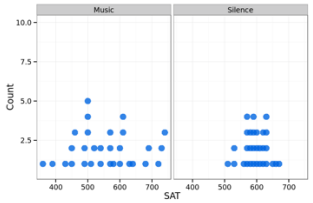
<table>
<tr><td colspan="2">Topic of Lesson:
Comparing two treatments from a randomized experiment</td><td>Date:</td></tr>
<tr><td colspan="3">Summary of Lesson:
This lesson centers on comparing two treatments from a randomized experiment. In this experiment, 60 hypothetical randomly selected students are randomly assigned to either a treatment group (music playing) or control group (silence). The main research question is “Does listening to music improve SAT scores?” Students are guided through the process of conducting a test to make inferences.</td></tr>
<tr><td>Prerequisites:
None</td><td>Activities:
N/A</td><td>Estimated Time:
Approximately 75 minutes (depends on the amount of time you want to spend implementing this lesson)</td></tr>
<tr><td colspan="2">Learning Objectives
Students will be able to …
• develop the need for a summary measure to compare two groups with quantitative data.
• determine if a single observed difference is meaningful or just due to chance.
• understand why we need to conduct a test to make a conclusion, rather than just looking at the observed difference.
• draw a conclusion about the experiment.</td><td>Assessments
• Optional HW assignment with problems of your choice (see comparing_treatments_hw_sample.docx file for possible problems to use)</td></tr>
<tr><td colspan="2">CCSSM Standards Addressed
• S.IC.5. Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.</td><td>CCSSM Mathematical Practices Emphasized
• MP1. Make sense of problems and persevere in solving them.
• MP4. Model with mathematics.
• MP5. Use appropriate tools strategically.</td></tr>
<tr><td colspan="3">Teaching Materials
• Varies (your decision) – graphing calculators, computers, tables, handouts, etc.
• Datasets for lesson: MusicvsSilence1.csv (meaningful difference), MusicvsSilence2.csv (not a meaningful difference)
• Datasets for homework: sleepPills.csv, treeGrowth.csv</td></tr>
</table>

Learning Plan for Lesson 5

Teacher Activity	Estimated Time
1. *Introduce the Experiment* To explore the effect of music on studying, researchers are interested in the following question *"Does listening to music effect reading comprehension?"* To investigate this question of interest, researchers conducted a randomized experiment to compare the effects of music on reading comprehension scores. 60 rural high school students in the U.S were randomly selected to be in the study. These students were then randomly assigned to two groups: 30 students were in a treatment group where music was played while they took a reading comprehension SAT test and 30 students were in a control group where no music was played while they took the same reading comprehension SAT test. The researchers recorded each student's score on the SAT test (scores on this test can range from 200 – 800 in increments of 10 points).	~ 5 min
2. *The Data* There are two datasets provided and one (or both) can be used throughout this lesson. When using a test to analyze the difference in means between the two groups (music and silence), ➢ **Dataset #1** (*MusicvsSilence1.csv*) has a meaningful difference. In other words, a test returns a small p-value (reject the null) indicating that there is a meaningful difference between the two groups. Therefore, with this dataset, we can conclude that listening to music effects reading comprehension scores. In fact, listening to music caused the SAT scores to decrease. ➢ **Dataset #2** (*MusicvsSilence2.csv*) does not have a meaningful difference. In other words, a test returns a large p-value (fail to reject the null) indicating that there is not a meaningful difference between the two groups. The mean SAT score does not differ for the treatment and control groups. Therefore, with this dataset, we can conclude that listening to music has no effect on reading comprehension scores. Both datasets are summarized below and boxplots showing how the means of each group compare are displayed. **Dataset #1** (*MusicvsSilence1.csv*)	~ 10 min (Depends on how the data is used)

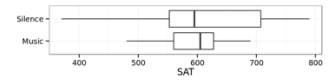


	Music	Silence
mean	555.67	596.33
sd	98.84	37.28
min	360.00	510.00
Q1	492.50	572.50
median	555.00	595.00
Q3	610.00	620.00
max	740.00	670.00
n	30.00	30.00

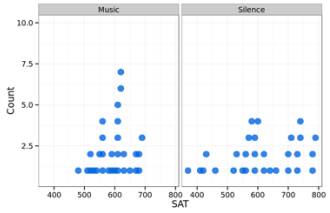


	SAT
Difference in Means	-40.67

Dataset #2 (*MusicvsSilence2.csv*)



	Music	Silence
mean	594.67	609.00
sd	55.32	117.16
min	480.00	370.00
Q1	560.00	552.50
median	605.00	595.00
Q3	627.50	707.50
max	690.00	790.00
n	30.00	30.00



	SAT
Difference in Means	-14.33

3. *Test for Difference Between Two Means*

Show students how to perform a test for difference between two means using one (or both) of the datasets and introduce vocabulary terms at your discretion. These tests can be implemented a variety of ways (graphing calculators, tables, web applets, GeoGebra, etc.)

The same web applet (http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/) can be used for comparing two treatments. Under One of Each, select Test or Estimate. The datasets described above have been pre-loaded in this web applet. To access these datasets, select Pre-Loaded Data from the drop down menu under How would you like to input the data? and then select the appropriate dataset under Available Datasets.

Note: **MusicvsSilence1** is Dataset #1 and **MusicvsSilence2** is Dataset #2.

Click on Use These Data and go through the process and steps of the test at your discretion. Below are suggestions to implement when comparing two treatments.

a) Name the test (and check assumptions)

Examples: Two-Sample t-test, randomization test.

b) State the hypotheses

In statistical language the idea of "no difference between treatment groups" is called the null hypothesis and is written in terms of the parameter. In other words, the two population means are equal or the difference between the means is zero. In this case, we are comparing two treatments and would write our null hypothesis as

$$H_0: \mu_1 - \mu_2 = 0 \text{ or } H_0: \mu_1 = \mu_2$$

where Group 1 represents the Music group (treatment) and Group 2 represents the Silence group (control).

We also have an alternative hypothesis, labeled H_a or H_1, which tells us the "direction" the researchers would like to conclude is true. This information is usually provided in the research question. Since the research question is "Does listening to music effect reading comprehension?", the researchers are thinking there might be a difference between the two groups. However, they are NOT specifying whether they think the control group will score higher on the reading comprehension SAT test than the treatment group or vice versa. In other words, they are looking for a difference in either direction. Therefore, in this case, our alternative hypothesis is

$$H_a: \mu_1 - \mu_2 \neq 0 \text{ or } H_a: \mu_1 \neq \mu_2$$

If the research question was stated as "Does listening to music improve reading comprehension?", then the researchers think that the Music group will score higher on the SAT test than the Silence group. As such, we would write the alternative hypothesis as

$$H_a: \mu_1 - \mu_2 > 0 \text{ or } H_a: \mu_1 > \mu_2$$

c) Compute the test statistic

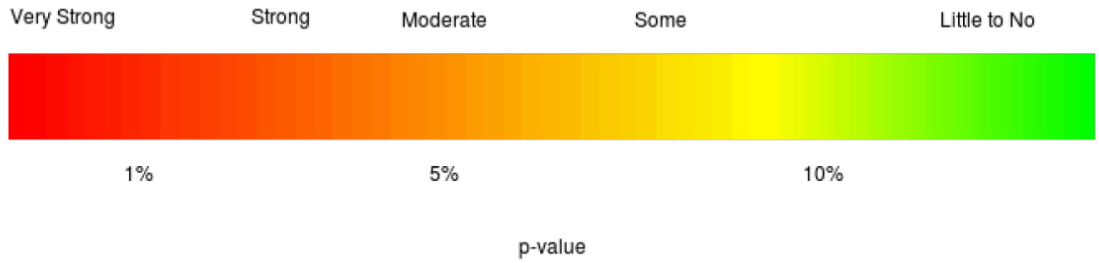
~ 55 min

(Depends on how the test is implemented and how many datasets are used)

First, calculate your observed result (i.e., the difference in population means: Group 1 (Music) mean – Group 2 (Control) mean).

We use the observed result to gauge the evidence against the null hypothesis. That's how the scientific method works. We formulate a hypothesis, which can be falsified, then see if the data collected argue against the hypothesis. Sometimes our observed result provides a lot of evidence against the null hypothesis – when the observed result is very unlikely under the null model – while other times it has very little evidence against the null hypothesis – when the observed result is likely under the null model. To explain to others how likely or unlikely the observed result is under the null model, we report the "strength of evidence" – also called the p-value.

To help interpret the strength of evidence, we offer the following picture. The important point is that smaller p-values (in red) provide stronger evidence against H_0. Larger p-values mean the data are not unusual under H_0.



Calculate a p-value using your method of choice.

d) State your conclusion

State a conclusion in the context of the problem. Did listening to music have an effect on reading comprehension?

Example: There is some evidence against the null hypothesis. Therefore, I would say that there is some evidence that listening to music has an effect on reading comprehension, but it is not super strong because a p-value of 0.079 is not really small.

4. *Wrap Up*

Take-Home Messages

- If there is no treatment effect, then the observed difference is just due to random chance (i.e., random assignment of treatments). In other words, there is no difference between the treatment groups.
- By randomly applying treatments, we are creating groups that should be very similar because differences between groups (e.g., SAT Prep course, IQ, etc.) are evened out by the random group allocation. If we see a difference between groups, then we doubt that treatments don't matter. Any difference between groups is caused by the treatment applied. Random assignment is a very powerful tool.

~ 5 min

Additional Take-Home Messages (if hypotheses and p-values were discussed)

- Setting up the null and alternative hypotheses is very important. The equals sign always goes into H_0. The direction in H_a must match the researchers' ideas – what they would like to show. It can be <, >, or ≠. The latter means they are looking for a difference in either direction.
- The idea of p-value is important in statistics. Stronger evidence means smaller p-values. Large p-values mean the data are not unusual under H_0.

5. *Homework* (OPTIONAL)

At your discretion, assign homework problems of your choice. For example, you could assign problems where students need to run through a test to draw a conclusion about an experiment.
The file, *comparing_treatments_hw_sample.docx*, contains sample problems and the solutions can be found in *comparing_treatments_hw_sample_solutions.docx*. These problems can be completed using the same web applet above (http://shiny.math.montana.edu/jimrc/IntroStatShinyApps/). However, the homework datasets have not been pre-loaded and students would need access to the two dataset (.csv) files so they can upload the datasets.

Under One of Each, select Test or Estimate. Select Local CSV File from the drop down menu under How would you like to input the data? and then choose and upload the appropriate file (sleepPills.csv or treeGrowth.csv). Make sure the box for (1) "Row One is column names" is selected, (2) the "Separator" selected is "Comma" and (3) select "None" under "Quote."



Finally, click on Use These Data. From here, the students can use the web applet to answer the homework questions.

Comparing Two Treatments Homework
Sample Problems

There are two datasets needed for these problems:
Problem 1 → *sleepPills.csv*
Problem 2 → *treeGrowth.csv*

1. We will be investigating the effectiveness of sleeping pills. Cushny and Peebles (1905) completed a study investigating the effects of two different soporific drugs (sleeping pills). They obtained 20 people to participate in the study and randomly assigned 10 to each drug. They measured the amount of extra sleep each person got on average while using the drug over a 5-day period.

   The research question you will be investigating is:

   "Is one drug more effective at increasing the number of hours slept?"

   a. Write the null and alternative hypotheses.

   b. What is the observed result (the difference in means: Drug 1 Mean – Drug 2 Mean) for this study?

   c. Obtain a p-value.

   d. Are you willing to conclude that one drug was more effective than the other? Explain.

2. How does exposure to ozone affect the growth of trees? Researchers asking this question randomly assigned small Sitka spruce trees to either grow in a regular growth chamber, or to grow in a chamber with increased levels of ozone. The trees were just the ones they had available to them, not sampled from all such trees.

   a. Identify the null hypothesis.
      i. The mean tree growth is the same in the regular and increased ozone environments in this sample.
      ii. Exposure to ozone has no effect on the true mean tree growth.
      iii. The true mean tree growth is higher in the ozone-enriched environment than in the regular environment.
      iv. There is an association between ozone levels and average tree growth.

   b. Write the alternative hypothesis.

   c. What is the observed result (the difference in means: Ozone Mean – Regular Mean) for this study?

   d. Obtain a p-value.

   e. Are you willing to conclude that ozone has an effect on tree growth? Explain.

Comparing Two Treatments Homework
SOLUTIONS
Sample Problems

1. We will be investigating the effectiveness of sleeping pills. Cushny and Peebles (1905) completed a study investigating the effects of two different soporific drugs (sleeping pills). They obtained 20 people to participate in the study and randomly assigned 10 to each drug. They measured the amount of extra sleep each person got on average while using the drug over a 5-day period.

   The research question you will be investigating is:

   "Is one drug more effective at increasing the number of hours slept?"

   a. Write the null and alternative hypotheses.

   $H_0: \mu_1 = \mu_2$ or $H_0: \mu_1 - \mu_2 = 0$
   $H_a: \mu_1 \neq \mu_2$ or $H_a: \mu_1 - \mu_2 \neq 0$

   where Group 1 = Drug 1 and Group 2 = Drug 2

   b. What is the observed result (the difference in means: Drug 1 Mean – Drug 2 Mean) for this study?

   0.75 – 2.33 = -1.58

   c. Obtain a p-value.

   p-value ≈ 0.079 = 7.9%

   d. Are you willing to conclude that one drug was more effective than the other? Explain.

   There is <u>some</u> evidence against the null hypothesis. Therefore, I would say that there is some evidence that one drug is more effective than the other, but it is not super strong because a p-value of 0.079 is not really small.

2. How does exposure to ozone affect the growth of trees? Researchers asking this question randomly assigned small Sitka spruce trees to either grow in a regular growth chamber, or to grow in a chamber with increased levels of ozone. The trees were just the ones they had available to them, not sampled from all such trees.

   a. Identify the null hypothesis.

   Correct Answer: ii

   b. Write the alternative hypothesis.

   $H_a: \mu_1 \neq \mu_2$ or $H_a: \mu_1 - \mu_2 \neq 0$

   where Group 1 = Ozone and Group 2 = Control.

   c. What is the observed result (the difference in means: Ozone Mean – Regular Mean) for this study?

   1.25 – 1.49 = -0.24

d. Obtain a p-value.

p-value ≈ 0.0056 = 0.56%

e. Are you willing to conclude that ozone has an effect on tree growth? Explain.

There is strong evidence against the null hypothesis. Therefore, I would say that ozone does have an effect on tree growth because a p-value of 0.0056 is not small.

Statistics Project

Description

Purpose:
The purpose of this project is to give you the opportunity to apply what you have learned throughout this unit. We have discussed normal distributions and normal probabilities, observational studies and experiments, interval estimation and margin of error, and comparing two treatments from a randomized experiment. You are to formulate a statistical question, decide how you want to answer it (via an observational study or an experiment), collect some data and analyze it, and draw a conclusion about your question.

Topics:
You may do your project on any topic. Make it interesting, appropriate, and have fun with it. For example, you could

- design an observational study and create a survey that asks people if they prefer to study with or without music playing.
- design an experiment to investigate whether students perform better on math questions while listening to music or no music.

Group Size:
You may work alone or with a partner for this project.

Proposal:
Submit a proposal about what you intend to study for this project. To get your project approved, you must be able to demonstrate how your study will meet the requirements of this project. In other words, you will need to

1. clearly and completely communicate your statistical question
    a. what is your question of interest?
2. fully describe the observational study or experiment
    a. what are you measuring?
    b. what is your response variable?
    c. how will you collect the data?
    d. who will be included in your sample?
    e. how are you going to select your sample?
    f. how will you analyze the data?

Make sure that your study will be safe and ethical if you are using human subjects (anonymous, able to quit at any time, permission to be in the study, etc.).

Poster:
The key to a good statistical poster is communication and organization. Create a poster that emphasizes all aspects of your project. It's a good idea to include the following in your poster:

- Title
- Question of interest
- Introduction to your study
- Data collection
- Any graphs or summary statistics you choose to include
- Data analysis
- Conclusion
- Pictures of your data collection

Presentation:
You will give a 3-5 minute oral presentation to the class about your poster and project.

Sample Exam Questions

Lesson 2 – Normal Distribution

1. When designing elevators, it is important to consider the weights of people so that elevators are not overloaded. According to the National Health Survey, the weight for adult makes in the U.S. is normally distributed with a mean weight of 173 pounds and a standard deviation of 30 pound (i.e., N(173, 30)).

    a. What proportion of U.S. adults weigh more than 203 pounds?

    b. What proportion of U.S. adults weigh between 143 and 233 pounds?

    c. What weight represents the 16^{th} percentile?

2. Suppose that the mean weight for men 18 to 24 years old is 170 pounds, and the standard deviation is 20 pounds. A man weighing 140 pounds lies
    a. 3 standard deviations below the mean.
    b. 3 standard deviations above the mean.
    c. 1.5 standard deviations below the mean.
    d. 1.5 standard deviations above the mean.

Lesson 3 – Observational Studies and Experiments

3. Determine whether the following study is an observational study or an experiment. Explain.

    Students in a small statistics course collected data to determine if the length of the forearm could be used to predict the length of the foot (both measured in centimeters) for all students at the university.

4. What's the primary difference between observational studies and experiments?
    a. Experiments must use a random sample.
    b. Observational studies will not have lurking variables.
    c. Experiments must use random assignment of participants to treatments.
    d. Experiments require assignment of participants to treatments.

5. 50 randomly selected athletes who had suffered sprained ankles were randomly assigned to one of two exercise programs. The first group returned to sports activities in a mean of 38.7 days. The other group returned to sports in a mean of 18.3 days. Why is it important to assign the athletes to the two treatments randomly?

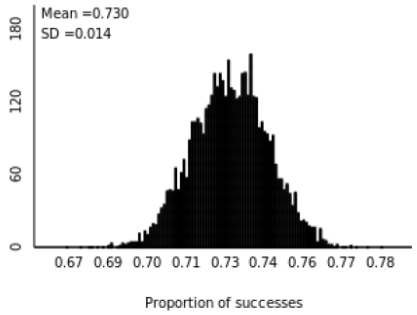
6. In 1984 a popular science magazine reported that a study of women in Italy found that having sons caused the lifespan of mothers to shorten by about 22 weeks per son, but that daughters would lengthen the mothers' lives. The data came from church records from 1975 – 1983. Is the article's claim valid?
    a. No. The data is too old to apply to the population in 1984.
    b. Yes. Since males are more stressful to raise than females, it makes sense that the mothers' lives would be shortened with sons.
    c. No. Causal inference cannot be drawn from an observational study.

7. True or False: Observational studies can establish cause-and-effect relationships?
    a. True
    b. False

Lesson 4 – Interval Estimation and Margin of Error

8. True of False: In the state of Alabama a random sample of 21 university students is taken. The sample proportion of Alabama citizens is 0.81. If we took another sample of 21 university students, the sample proportion will always be 0.81.
    a. True
    b. False

9. A poll of 975 randomly selected Americans asked if they preferred to watch a movie at home rather than in the theater. Based on the results, we will build a 95% confidence interval for the true proportion of U.S. adults who prefer watching movies at home. The following plot was obtained by simulation.



    a. We didn't say what proportion of the sample prefers to watch at home, but you can tell from the plot. What is the sample proportion?

    b. What is the margin of error for a 95% confidence interval?

    c. Using your answer above, construct the 95% confidence interval.

10. The American Management Association polled a sample of 304 U.S. businesses asking each if they monitor employee's website visits. Of these, 201 said they did. We will assume this is a representative sample of all U.S. business and we want to estimate the true proportion of businesses which monitor web activity. A simulation gives a standard error for the sample proportion of 0.027.

    a. Compute the margin of error for a 95% confidence interval.

    b. Using your answer above, construct the 95% confidence interval.

Lesson 5 – Comparing Two Treatments

11. Suppose you are hired by a local pizza company "C" and your job is to see if your company can make the advertising claim: "Our large size pizzas are larger in diameter than those of Dominoes."

    Write the null and alternative hypotheses.

12. Researchers conducted a study in which they invited nutrition experts to an ice cream social. These experts were randomly given either a 17- or a 34-ounce bowl. They were then invited to serve themselves ice cream. The suspicion was that even nutrition experts would tend to take more ice cream when given a larger bowl. Results for the actual volumes of ice cream taken (in ounces) are summarized in the following table. Assume the data are approximately bell-shaped and symmetric.

	Sample size	Sample mean	Sample SD
17-ounce bowl	20	3.38 oz	2.05 oz
34-ounce bowl	17	5.81 oz	2.26 oz

    a. Write the null and alternative hypotheses.

    b. What is the observed result (34-ounce Mean – 17-ounce Mean)?

    c. Calculate a p-value.

    d. State a conclusion.

Sample Exam Questions
SOLUTIONS

Lesson 2 – Normal Distribution

1. When designing elevators, it is important to consider the weights of people so that elevators are not overloaded. According to the National Health Survey, the weight for adult makes in the U.S. is normally distributed with a mean weight of 173 pounds and a standard deviation of 30 pound (i.e., N(173, 30)).

    a. What proportion of U.S. adults weigh more than 203 pounds? *0.16*

    b. What proportion of U.S. adults weigh between 143 and 233 pounds? *0.815*

    c. What weight represents the 16th percentile? *143*

2. Suppose that the mean weight for men 18 to 24 years old is 170 pounds, and the standard deviation is 20 pounds. A man weighing 140 pounds lies
    b. 3 standard deviations below the mean.
    c. 3 standard deviations above the mean.
    d. 1.5 standard deviations below the mean.

e. 1.5 standard deviations above the mean.

Correct Answer: c

Lesson 3 – Observational Studies and Experiments

3. Determine whether the following is an observational study or an experiment. Explain.

   Observational study because the researcher is not manipulating anything (i.e., not assigning any treatments). The researcher is simply recording the lengths of forearms and feet.

4. What's the primary difference between observational studies and experiments?

   Correct Answer: d

5. 50 randomly selected athletes who had suffered sprained ankles were randomly assigned to one of two exercise programs. The first group returned to sports activities in a mean of 38.7 days. The other group returned to sports in a mean of 18.3 days. Why is it important to assign the athletes to the two treatments randomly?

   In order to even out the lurking variables between the two groups so that causal inferences can be drawn.

6. In 1984 a popular science magazine reported that a study of women in Italy found that having sons caused the lifespan of mothers to shorten by about 22 weeks per son, but that daughters would lengthen the mothers' lives. The data came from church records from 1975 – 1983. Is the article's claim valid?

   Correct Answer: c

7. True or False: Observational studies can establish cause-and-effect relationships?

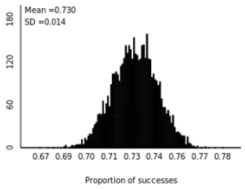
   Correct Answer: False

Lesson 4 – Interval Estimation and Margin of Error

8. True of False: In the state of Alabama a random sample of 21 university students is taken. The sample proportion of Alabama citizens is 0.81. If we took another sample of 21 university students, the sample proportion will always be 0.81.

   Correct Answer: False

9. A poll of 975 randomly selected Americans asked if they preferred to watch a movie at home rather than in the theater. Based on the results, we will build a 95% confidence interval for the true proportion of U.S. adults who prefer watching movies at home. The following plot was obtained by simulation.



a. We didn't say what proportion of the sample prefers to watch at home, but you can tell from the plot. What is the sample proportion? *0.730*

b. What is the margin of error for a 95% confidence interval?

ME = 2×SE = 2×0.014 = 0.028

c. Using your answer above, construct the 95% confidence interval.

$0.730 \pm 2\times SE = 0.730 \pm 0.028 = (0.702, 0.758)$

10. The American Management Association polled a sample of 304 U.S. businesses asking each if they monitor employee's web site visits. Of these, 201 said they did. We will assume this is a representative sample of all U.S. business and we want to estimate the true proportion of businesses which monitor web activity. A simulation gives a standard error for the sample proportion of 0.027.

a. Compute the margin of error for a 95% confidence interval.

ME = 2×SE = 2×0.027 = 0.054

b. Using your answer above, construct the 95% confidence interval.

$\frac{201}{304} \pm 2\times SE = 0.661 \pm 0.054 = (0.607, 0.715)$

Lesson 5 – Comparing Two Treatments

11. Suppose you are hired by a local pizza company "C" and your job is to see if your company can make the advertising claim: "Our large size pizzas are larger in diameter than those of Dominoes."

Write the null and alternative hypotheses.

$H_0: \mu_C - \mu_D = 0$ *or* $H_0: \mu_C = \mu_D$
$H_a: \mu_c - \mu_D > 0$ *or* $H_a: \mu_C > \mu_D$

where C = Local pizza company and D = Dominoes

12. Researchers conducted a study in which they invited nutrition experts to an ice cream social. These experts were randomly given either a 17- or a 34-ounce bowl. They were then invited to serve themselves ice cream. The suspicion was that even nutrition experts would tend to take more ice cream when given a larger bowl. Results for the actual volumes of ice cream taken (in ounces) are summarized in the following table. Assume the data are approximately bell-shaped and symmetric.

	Sample size	Sample mean	Sample SD
17-ounce bowl	20	3.38 oz	2.05 oz
34-ounce bowl	17	5.81 oz	2.26 oz

e. Write the null and alternative hypotheses.

$H_0: \mu_1 - \mu_2 = 0$ *or* $H_0: \mu_1 = \mu_2$
$H_a: \mu_1 - \mu_2 > 0$ *or* $H_a: \mu_1 > \mu_2$

where Group 1 = 34-ounce bowl and Group 2 = 17-ounce bowl

f. What is the observed result (34-ounce Mean – 17-ounce Mean)?

5.81 – 3.38 = 2.43

g. Calculate a p-value.
p-value ≈ 0.0018

h. State a conclusion.

There is strong evidence against the null hypothesis. Therefore, I would say that nutrition experts take more ice cream when given a larger bowl.

APPENDIX B

CONSENT FORM

PARTICIPANT CONSENT FORM
FOR
PARTICIPATION IN HUMAN RESEARCH AT
MONTANA STATE UNIVERSITY

Investigating the Teaching of Statistics at the High School Level: Algebra II and Mathematics III

What this study is about
Because you teach high school mathematics (particularly, Algebra II or Mathematics III), you are being asked to participate in a research study about the teaching of Statistics in these courses at the high school level. This research will take place during the 2015 – 2016 academic school year, and will help researchers better understand how to support high school teachers to teach Statistics that aligns with Common Core State Standards.

What you will be asked to do
There are three parts to this research study, and your participation is voluntary. If you agree to participate and are selected to be included in this research study, you are agreeing to the following:

- Phase 1: Complete two surveys that help the researcher to understand about you as a Mathematics and Statistics teacher. Together, these surveys are expected to take no longer than 90 minutes and will be completed by June 30, 2015. By September 30, 2015, you will be asked to complete an additional survey that will take approximately 20 minutes.

- Phase 2: Agree to use five Statistics-based lesson plans within a Statistics unit that will be provided to you. If you agree, the researcher will observe you teaching these lessons. The unit and lessons are based on the Statistics content standards that are recommended by the Common Core State Standards Mathematics to be taught in Algebra II or Mathematics III. These classroom observations will take place at a time that is most convenient for you. If the researcher is unable to complete the observation in your classroom, she is able to use Skype or video recordings of your class sessions instead. Additionally, you may be asked to answer brief follow-up questions after each teaching observation.

- Phase 3: Complete an exit survey expected to take no longer than 60 minutes. Additionally, you may be asked to participate in a recorded interview lasting approximately 30 minutes. This interview will provide the researcher with extra information regarding the teaching of the five Statistics lessons.

Risks and benefits
Risks for participating in this study are minimal. You might feel uncomfortable when you are answering survey questions or when you are being observed. During observations, if video recording is necessary, the camera will be focused only on the front of the classroom to reduce interaction with and video recording of any students. At the end of the study, the researcher will provide all participants with the Statistics unit that was developed and refined as part of the study.

Your participation is voluntary
Your participation in this research study is voluntary, and your participation (or non-participation) will not affect your professional standing. You are free to withdraw your consent and discontinue participation, stop a survey, assessment, observation, or interview, or refuse to answer any question at any time. If you choose to participate, it will not affect your right to receive the Statistics unit at the end of the study.

Information is confidential
To prevent your data from being associated with you personally, you have been assigned a unique 3-digit identification code that has been emailed to you for use with this and all future surveys. Your name will not appear anywhere in your responses to this survey. Use of your ID code will ensure that we can consistently track the data you provide across the entire project, without your name being stored in the same place as your data. The code and all other data will be stored on secure computers and databases. The risk that your confidential data from interviews or surveys will be seen by someone who does not have permission is minimal, and we are taking every precaution to prevent that from occurring.

Information you enter in the following survey is confidential. Only the researcher, Elizabeth Arnold, will have access to the computer and databases on which project information will be stored. No individual participant's responses to this survey will be shared with anyone outside of the research project. Data will be reported in the aggregate, and no identifying information will be published in any report.



Contact information
If you have questions about this research project, you may contact Elizabeth Arnold, (406) 994-5348 [liz.g.arnold@gmail.com]. If you have questions regarding your rights as a participant, any concerns regarding this project or any dissatisfaction with any aspect of this study, you may report them – confidentially, if you wish – to the Chair of the Institutional Review Board, Mark Quinn, (406) 994-4707 [mquinn@montana.edu].

By completing any survey or assessment or participating in any observation or interview, you are granting your consent to participate in this research. Type your name in this box indicating your consent.

APPENDIX C

ADJUSTED TPACK RUBRIC

Adapted TPACK Rubric

	TPACK Levels			
TPACK Components	*Recognizing (1)*	*Accepting (2)*	*Adapting (3)*	*Exploring/Advancing (4)*
1. Purpose for Technology *An overarching conception about the purpose of incorporating technology in teaching statistics topics.*	• Cognitive technology is used for motivation rather than actual subject matter development. *(e.g., playing a video or calculating a mean. Here, technology is used to motivate the lesson or showcase the efficiency of technology versus by-hand calculations.)* All learning of new ideas presented by the teacher mostly without technology; students are not actively involved. *(Technology is being used just for the sake of using it.)* • Technology-based activities do not include inquiry tasks. *(e.g., step-by-step, detailed instructions are provided.)* Technology procedures concentrate on drills and practice only.	• Cognitive technology is used for motivation rather than actual subject matter development. *(e.g., playing a video or calculating a mean. Here, technology is used to motivate the lesson or showcase the efficiency of technology versus by-hand calculations.)* Larger part of technology use (by the teacher) is for demonstrations, which include presenting new knowledge. • Technology-based activities do not include inquiry tasks. Technology procedures concentrate on teacher demonstrations and practice. *(e.g., teacher demonstrates technology and students practice. Here, students may or may not follow along and participate.)*	• Teacher is the one using cognitive technology in a way that is new and different from teaching without this technology. A dynamic use of technology is employed. *(e.g., demonstrating a simulation, removing a data point to see instant changes.)* • Technology-based activities do include inquiry tasks, but these tasks are not of high cognitive demand. Students are not doing statistics. The context and interpretation of problems are not included.	• Students are the primary users of cognitive technology. The technology allows students to explore and experiment with it for new knowledge and the opportunity to deepen conceptual understanding of statistical concepts. • Technology-based activities do include inquiry tasks that are of high cognitive demand. Students are doing statistics. The context and interpretation of problems are included.
2. Students' Understandings *Knowledge of students' understandings, thinking, and learning statistics with technology.*	• Cognitive technology is used primarily for student practice. • Digital materials present space allocated solely for applications and drills.	• Cognitive technology is primarily used for teacher demonstrations or teacher-led "follow along" work with technology. If used, it is rarely used for students' independent explorations.	• Teacher focuses on students' thinking of statistics while students are using the cognitive technology on their own – both for learning new knowledge and review of prior knowledge. *(e.g., students use technology to*	• Teacher acts as a guide, not a director, and promotes students' statistical conceptual understanding and thinking with cognitive technology.

		• Digital materials mirror the structure of the textbook presentation of statistics. They do not include active explorations.	*reinforce previously teacher-taught concepts)* • Digital materials provide an environment for students to participate in their learning of statistics with <u>teacher guidance</u>.	• Digital materials provide an environment for students to do statistics and make meaningful connections. *(A little teacher guidance might be necessary to help students make meaningful connections.)*
3. <u>Curriculum</u> *Knowledge of curriculum and curricular materials that integrate technology in learning and teaching statistics.*	• Teacher does not use cognitive technology for <u>learning</u> statistics. *(Technology is present, but not focused on content.)* • Cognitive technology, if used, is not aligned with one or more of the Statistics Unit's learning goals.	• Teacher uses the standard approach to the statistics topics with cognitive technology being used as an add-on. *(Use of technology is an afterthought.)* • Cognitive technology is <u>partially aligned</u> with one or more of the Statistics Unit's learning goals. *(Teacher demonstrates difficulty in identifying statistics topics for including technology as a tool for learning.)*	• Cognitive technology is used as a replacement for non-technology based tasks in a traditional formula-based approach for the teaching and learning of statistics. *(e.g., using GeoGebra to find normal probabilities)* • Cognitive technology is <u>aligned</u> with one or more of the Statistics Unit's learning goals. However, technology is not used for students' independent thinking and exploration. *(Technology is not always appropriate for the statistics topic.)*	• Teacher uses cognitive technology in a constructive way, including tasks for the development of higher level thinking and deepening understanding of statistics topics. *(Students are given tasks with technology to expand their understanding.)* The teacher challenges the traditional formula-based approach for the teaching and learning of statistics. *(e.g., using simulations to construct and interpret a margin of error)* Students are engaged in learning statistics with technology. • Cognitive technology is aligned with the Statistics Unit's learning goals. Technology use is effective and appropriate. *(Technology is used students' independent thinking and exploration.)*

4. <u>*Instructional Strategies*</u> *Knowledge of instructional strategies and representations for teaching and learning statistics with technology.*	• Teacher focuses on how to use cognitive technology rather than how to explore. *(e.g., step-by-step instructions to calculate a summary statistic)* Teacher-directed lectures followed by student practice. • Digital materials provide students only with opportunities for drill and practice.	• Approach to instruction is teacher-led. Teacher structures the lesson plan with limited student explorations with cognitive technology. • Digital materials are <u>not</u> built around the Statistics Unit's learning goals and do <u>not</u> promote student reflection *(Students do not need to engage with the material. There is no learning occurring. An activity is used just for the sake of using an activity)*	• Approach to instruction is teacher-led. There is some student exploration, but the teacher controls the tempo to maintain the progression of activities. • Digital materials are built around the Statistics Unit's learning goals, but do <u>not</u> fully promote student reflection – especially the posing of questions for sense making. *(There is some reflection, but students are not reflecting in a way that promotes conceptual understanding of the statistics topic.)*	• Approach to instruction is primarily student-led. There may be some teacher-led instruction and strategies intertwined in the lesson to focus on students' thinking and conceptual understanding of statistics. Teacher's use of cognitive technology is beyond the traditional formula-based approach to statistics topics. • Digital materials are built around the Statistics Unit's learning goals, and explicitly and fully promote student reflections – especially the posing of questions for sense making *(Students justify their answers.)*

APPENDIX D

SURVEYS

Enrollment Survey

Dear Teacher:

The following survey will collect some basic contact information from you so we can build a project database of participants. Your address, phone number, and e-mail addresses will not be used except for official project business and will not be shared with anyone outside of the project. IMPORTANT: Your responses will be imported into our database exactly the way you type them. Please capitalize and punctuate as you would if addressing an official letter to yourself.

We also will ask you a few questions about the classes you are teaching during the 2015-2016 academic year. All questions are required unless otherwise stated. If your screen does not advance when you click "Save and Continue," look for a red error message above one or more questions that indicates a missing answer. Thank you for your time!

Let's get started:

1. Please enter your name, last name first:

2. So that we may contact you:
    a. Please enter your work e-mail address:
    b. Enter your work-email address again:

3. Please specify the state you will be teaching in during the 2015-2016 academic school year.

4. During the 2015-2016 academic school year, will you be teaching at least one section of Algebra II, Mathematics III, or the equivalent of these courses?
    - ☐ Yes - Algebra II or equivalent course
    - ☐ Yes - Mathematics III or equivalent course
    - ☐ No
    - ☐ Other (please specify)

OPTIONAL: We may, on occasion, need to contact you outside of school hours as the study progresses. The information you provide below will remain confidential.

5. Please provide your home mailing address. (It will help us if you capitalize as you would if addressing an official letter. Please use postal abbreviations CA, MT, or VA.)
    a. Street Address:
    b. Town:
    c. State:
    d. Zip Code:

6. It's important that we can reach you by e-mail during the summer to give you details relating to the study. Please provide an e-mail address that you check *regularly* during the summer months.
    a. Please enter your summer e-mail address:
    b. Enter your summer e-mail address again:

7. On a rare occasion we may need to contact you by phone during the summer: for example, if we have trouble reaching you by e-mail. Please provide a home or mobile phone number:

You have reached the end of the Teacher Enrollment Survey! You may navigate back to modify any answers if you wish. If you are finished, please click "SUBMIT." Once you click "SUBMIT," you will be

transferred to Montana State University's Math Department home page, signifying that we have received your responses, and you will not be able to return to the survey.

Initial Survey

Dear Teacher:

Before moving to the first page of this survey, please read the following explanation of how we safeguard your confidential responses. Entering your name in the box below and clicking "Save and Continue" will signify that you have read this policy and understand its protections.

Statement of Informed Consent

The first part of this survey is designed to capture current information about your workplace, technology accessible to you, and your overall willingness to teach from a provided statistics unit. Questions are required unless labels otherwise. If your screen does not advance when you click "Save and Continue," look for a red error message above one or more questions that indicates a missing answer. Thank you for your time!

Let's get started:

1. Enter the 3-digit ID code sent to you by e-mail (Please enter this ID *very carefully*):

2. Write the name of the school district you will be teaching in during the 2015-2016 academic year.

3. Write the name of the school you will be teaching in during the 2015-2016 academic school year.

4. List the classes (and times) you will be teaching during the 2015-2016 academic school year.

5. Will you have access to a class set of tablets or computers that have access to the Internet? If so, please specify what you have access to (e.g., iPad 2) in the comment box.
   - ☐ No
   - ☐ Yes, and I have access to _____
6. Will you have access to a class set of scientific graphing calculators? Is so, specify what you have access to (e.g., TI-83) in the comment box.
   - ☐ No
   - ☐ Yes, and I have access to _____

7. Do you have flexibility around when you teach a statistics unit in Algebra II or Mathematics III? For instance, are you required to teach it in November or are you able to teach it when it fits your schedule?

8. On a scale of 1 to 5 (Not Willing to Willing), state your willingness to use detailed lesson plans that involve the use of technology and inquiry-based activities for student explorations in statistics.

The next part of this survey is designed to capture your current thinking about the use of technology and inquiry based activities in your classroom. Questions are required unless labels otherwise. If your screen does not advance when you click "Save and Continue," look for a red error message above one or more questions that indicates a missing answer.

9. Describe the role of exploration and inquiry in the teaching and learning of mathematics in your classroom.

10. If you have used inquiry and exploration tasks in your mathematics classroom, describe the activity and how it was used in the classroom. If you have not used activities that involved inquiry tasks and exploration, please explain why.

11. Describe the role of using technology to teach mathematics in your classroom.

12. If you have used technology in your mathematics classroom in the past 6 months, describe *how* you have used it. Focus on what you believe to be the most productive use of technology. If you have not used technology in teaching mathematics, please explain why.

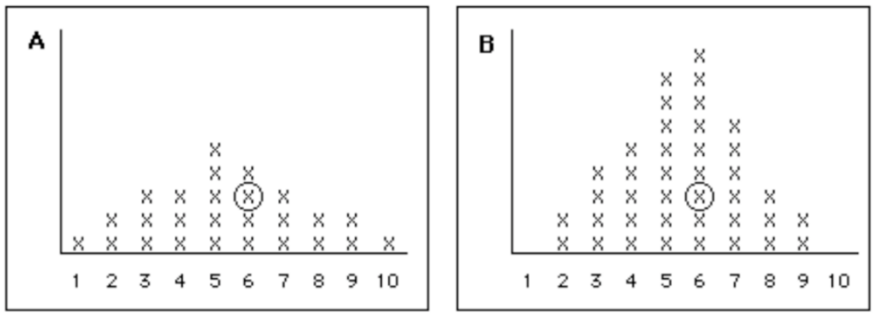
13. If your students have used technology in your mathematics classroom in the past 6 months, describe *how* they have used it. Focus on what you believe to be the most productive use of technology.

14. Briefly describe your role as a teacher when using technology to teach mathematics in your classroom. *How* are you using it in your role as a teacher?

15. Of the times that you have used technology to teach mathematics in your classroom this academic year, approximately what percent of the time are the students rather than the teacher the ones who are primarily using the technology?

The final part of this survey is designed to capture your current statistical content knowledge about Sampling Variability. Questions are required unless labels otherwise. If your screen does not advance when you click "Save and Continue," look for a red error message above one or more questions that indicates a missing answer.

16. Figure A represents the weights for a sample of 26 pebbles, each weighed to the nearest gram. Figure B represents the mean weights of a random sample of 3 pebbles each, with the mean weights rounded to the nearest gram. One value is circled in each distribution.



Is there a difference between what is represented by the X circled in A and the X circled in B? Please select the best answer from the list below.

- ☐ No, in both Figure A and Figure B, the X represents one pebble that weighs 6 grams.
- ☐ Yes, Figure A has a larger range of values than Figure B.
- ☐ Yes, the X in figure A is the weight for a single pebble, while the X in figure B represents the average weight of 3 pebbles.

17. In a geology course, students were learning to use a balance scale to make accurate weighings of rock samples. One student plans to weigh a rock 20 times and then calculate the average of the 20 measurements to estimate her rock's true weight. A second student plans to weigh a rock 5 times and calculate the average of the 5 measurements to estimate his rock's true weight. Which student is more likely to come the closest to the true weight of the rock he or she is weighing?

    ☐ The student who weighed the rock 20 times.
    ☐ The student who weighed the rock 5 times.
    ☐ Both averages would be equally close to the true weight.

18. Suppose half of all newborns are girls and half are boys. Hospital A, a large city hospital, records an average of 50 births a day. Hospital B, a small, rural hospital, records an average of 10 births a day. On a particular day, which hospital is less likely to record 80% or more female births?

    ☐ Hospital A (with 50 births a day), because the more births you see, the closer the proportions will be to .5.
    ☐ Hospital B (with 10 births a day), because with fewer births there will be less variability.
    ☐ The two hospitals are equally likely to record such an event, because the probability of a boy does not depend on the number of births.

19. A random sample of 25 college statistics textbook prices is obtained and the mean price is computed. To determine the probability of finding a more extreme mean than the one obtained from this random sample, you would need to refer to:

    i. the population distribution of all college statistics textbook prices.
    ii. the distribution of prices for this sample of college statistics textbooks.
    iii. the sampling distribution of textbook prices for all samples of 25 textbooks from this population.

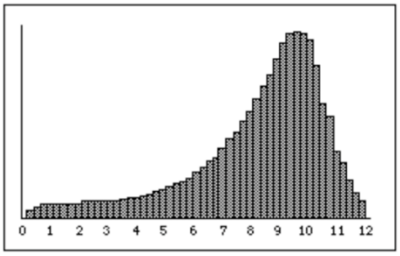
20. Consider the distribution of average numbers of hours that college students spend sleeping each weeknight. This distribution is very skewed to the right, with a mean of 5 and a standard deviation of 1. A researcher plans to take a simple random sample of 18 college students. If we were to imagine that we could take all possible random samples of size 18 from the population of college students, the sample distribution of average numbers of hours spent sleeping will have a shape that is

    i. Exactly normal.
    ii. Less skewed than the population.
    iii. Just like the population (i.e., very skewed to the right).
    iv. It's impossible to predict the shape of the sampling distribution.

21. Imagine you have a huge jar of candies that are a generic version of M&Ms. We know that 40% of the candies in the jar are brown. Imagine that you create a sample by randomly pulling 20 candies out of the jar. If you repeated this 10 times to create 10 samples, each with 20 candies, about how many browns would you expect to find in each of the 10 samples?

    i. Each sample would have exactly 8 brown candies.
    ii. Most of the samples would have 0 to 8 brown candies.
    iii. Most of the samples would have 8 to 20 brown candies.
    iv. Most of the samples would have 6 to 10 brown candies.
    v. You are just as likely to get any count of brown candies between 0 and 20.

Items 7 and 8 refer to the following situation:
The distribution for a population of measurements is presented below.



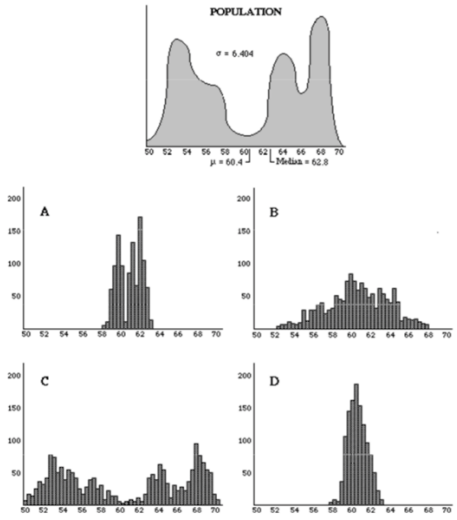
22. A sample of 10 randomly selected values will be taken from the population and the sample mean will be calculated. Which of the following intervals is MOST likely to include the sample mean?

    i. 4 to 6
    ii. 7 to 9
    iii. 10 to 12

23. Another sample of 10 randomly selected values will be taken from the population and the sample mean will be calculated. Which of the following intervals is LEAST likely to include the sample mean?

    ☐ 0 to 3
    ☐ 4 to 7
    ☐ 8 to 11

Items 9 to 14 refer to the following situation:
A hypothetical distribution for a population of test scores is displayed below. The population has a mean of 60.4, a median of 62.8, and a standard deviation of 6.404. Each of the other four graphs labeled A to D represent possible distributions of sample means for random samples drawn from the population.



24. Which graph best represents a distribution of sample means for 1000 samples of size 4?

    i. A
    ii. B

iii. C
iv. D

25. What do you expect for the shape of the sampling distribution (the distribution of sample means for all possible samples of size n=4)?

    i. Shaped more like a normal distribution than like the population distribution.
    ii. Shaped more like the population distribution than like a normal distribution
    iii. Shaped like neither the population nor the normal distribution.

26. What do you expect for the variability (spread) of the sampling distribution?

    i. Same as the population.
    ii. Less variability than the population (a narrower distribution).
    iii. More variability than the population (a wider distribution).

27. Which graph best represents a distribution of sample means for 1000 samples of size 50?

    i. A
    ii. B
    iii. C
    iv. D

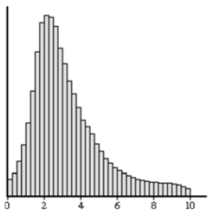
28. What do you expect for the shape of the sampling distribution (the distribution of sample means for all possible samples of size n=50)?

    i. Shaped more like a normal distribution.
    ii. Shaped more like the population.
    iii. Shaped like neither the population nor the normal distribution.

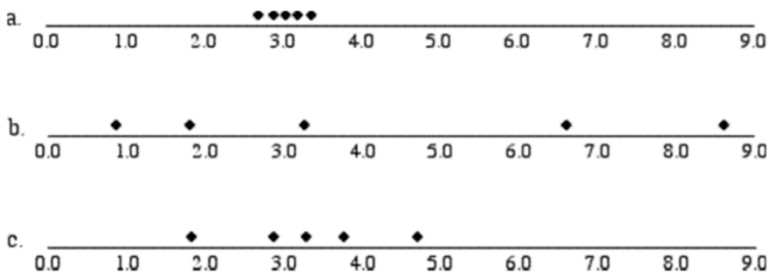
29. What do you expect for the variability (spread) of the sampling distribution?

    i. Same as the population.
    ii. Less variability than the population (a narrower distribution).
    iii. More variability than the population (a wider distribution).

The distribution for a population of measurements is presented below. The mean is 3.2 and the standard deviation is2. Suppose that five students each take a sample of ten values from the population and each student calculates the sample mean for his or her ten data values. The students draw a dotplot of their five sample means on the classroom board so that they can compare them.



30. Which of the following dotplots do you think is the most plausible for the one they drew on the board?



☐ A
☐ B
☐ C

Submit Your Answers

1. Please re-enter 3-digit ID code sent to you by email (Please enter this ID *very carefully*).

You have reached the end of the survey. You may navigate back to modify any answers if you wish. If you are finished, please click "SUBMIT." Once you click "SUBMIT," you will be transferred to Montana State University's Math Department home page, signifying that we have received your responses, and you will not be able to return to the survey.

Exit Survey

Dear Teacher:

The purpose of this survey is to help me understand your *current* thinking about the use of technology and inquiry in your mathematics classroom. At the beginning of this research study, you answered **7** open-ended questions meant to capture your thinking about the use of technology and inquiry-based activities in your mathematics classroom. These same **7** questions are presented below with an additional question regarding the use of simulations and a space to provide any other information or comments that you would like to share. Feel free to be as concise or reflective as you deem appropriate in your responses.

Thank you for your time completing this survey. Please email me at liz.g.arnold@gmail.com if you have any questions or concerns.

1. Enter the 3-digit ID code sent to you by e-mail (Please enter this ID very carefully):

Inquiry Questions

2. Describe the role of exploration and inquiry in the teaching and learning of mathematics in your classroom.

3. If you have used inquiry and exploration tasks in your mathematics classroom, describe the activity and how it was used in the classroom. If you have not used activities that involved inquiry tasks and exploration, please explain why.

Technology Questions

4. Describe the role of using technology to teach mathematics in your classroom.

5. If you have used technology in your mathematics classroom in the past 6 months, describe *how* you have used it. Focus on what you believe to be the most productive use of technology. If you have not used technology in teaching mathematics, please explain why.

6. If your students have used technology in your mathematics classroom in the past 6 months, describe *how* they have used it. Focus on what you believe to be the most productive use of technology.

7. Briefly describe your role as a teacher when using technology to teach mathematics in your classroom. *How* are you using it in your role as a teacher?

8. Of the times that you have used technology to teach mathematics in your classroom this academic year, approximately what percent of the time are the students rather than the teacher the ones who are primarily using the technology?

Randomization and Simulations Question

9. The activities in Lessons 3 (Observational Studies vs. Experiments) and 4 (Interval Estimation & Margin of Error) used a web applet that made use of randomizations and simulations. Do you think the integration of randomizations and simulations for learning statistical concepts is a good idea? Please explain why or why not and briefly describe any value you found in using the web applet.

Additional Comments

10. Please provide any additional comments that you would like to share about the integration of technology, inquiry-based activities and/or use of simulations in your mathematics classroom.

Thank you for taking time to answer these questions!

Teacher Background and Experience Survey

Dear Teacher:

The purpose of this survey is to collect some information about your professional and educational background and experience, as well as some demographic information about your class and school. There are a total of 9 questions with an additional question to provide any other information or comments that you would like to share.

Your responses will help strengthen my research by allowing me to characterize the pool of teachers in this study. This will ultimately allow me to better communicate the results of my study and broaden the scope of this research.

Thank you for your time in filling out this survey. Please email me at liz.g.arnold@gmail.com if you have any questions or concerns.

Let's get started:

1. Enter the 3-digit ID code sent to you by email (Please enter this ID very carefully):

Statistics and Technology Background Questions

2. Did you complete a Statistics course in your collegiate study? If possible, describe the highest level Statistics course you took (e.g., Elementary Statistics, Intermediate Statistics).

3. How often have you participated in professional development programs, conferences, or workshops that emphasize the integration of ***technology*** in a math classroom and/or the ***teaching of statistics***?

4. Briefly describe (if applicable) what things and/or concepts you have learned about ***integrating technology*** or ***teaching statistics*** during these programs/conferences/workshops.

Education Background Questions

5. What is the highest degree that you hold?
    a. BA or BS
    b. MA, MS, or Med
    c. Multiple MA, MS, or Med
    d. Phd or EdD
    e. Other (please specify)
6. What was your major field of study for the bachelor's degree?
    a. Elementary Education
    b. Secondary Education: Mathematics
    c. Secondary Education: Mathematics-Intensive field (e.g., engineering, statistics, physics)
    d. Secondary Education: Other field
    e. Mathematics
    f. Other Mathematics-Intensive field (e.g., engineering, statistics, physics)
    g. Other (please specify)

7. Do you have a minor, additional major, or a graduate degree?
    a. I don't have a minor, additional major, or a graduate degree.
    b. I have a minor in _____.
    c. I have an additional major in _____.
    d. I have a graduate degree in _____.

Teaching Experience and School Information Questions

8. Including this year, and rounding up to a whole numeral, how many years have you taught mathematics within grades 8 - 12?

9. Including this year, and rounding up to a whole numeral, how many times have you taught the following courses?
    a. Algebra II, Math III, or the equivalent?
    b. AP Statistics?
    c. General Statistics?

10. Regarding the school you teach at ...
    a. what is the approximate number of students in the entire school?

b. what is the approximate number of students in your Algebra II or Mathematics III (or the equivalent) class(es)?
c. what is the approximate number of teachers in your school who teach at least one class of mathematics or statistics?

Additional Comments

11. Please provide any other characteristics or information you would like to share.

Thank you again for taking time out of your day to complete this survey!

APPENDIX E

OBSERVATION PROTOCOL

Observation Protocol

Lesson:
Date:
Participant:
Time:
Technology Used:

Implementation	**Struggles with Content**
	Struggles with Technology
	Interaction with Students
	Other