

Illinois State University

ISU ReD: Research and eData

---

Theses and Dissertations

---

2013

## Examining Middle School Students' Statistical Thinking while Working in a Technological Environment

Melissa Arnold Scranton

Illinois State University, [missy.scranton@mchsi.com](mailto:missy.scranton@mchsi.com)

Follow this and additional works at: <https://ir.library.illinoisstate.edu/etd>



Part of the [Elementary and Middle and Secondary Education Administration Commons](#), [Junior High, Intermediate, Middle School Education and Teaching Commons](#), [Mathematics Commons](#), and the [Science and Mathematics Education Commons](#)

---

### Recommended Citation

Scranton, Melissa Arnold, "Examining Middle School Students' Statistical Thinking while Working in a Technological Environment" (2013). *Theses and Dissertations*. 44.

<https://ir.library.illinoisstate.edu/etd/44>

This Dissertation-Open Access is brought to you for free and open access by ISU ReD: Research and eData. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ISU ReD: Research and eData. For more information, please contact [ISURed@ilstu.edu](mailto:ISURed@ilstu.edu).

EXAMINING MIDDLE SCHOOL STUDENTS' STATISTICAL THINKING  
WHILE WORKING IN A TECHNOLOGICAL ENVIRONMENT

Melissa Arnold Scranton

248 Pages

December 2013

The purpose of this study was to gain a better understanding of how students think in a technological environment. This was accomplished by exploring the differences in the thinking of students while they worked in a technological environment and comparing this to their work in a paper and pencil environment. The software program *TinkerPlots: Dynamic Data Exploration* (Konold & Miller, 2005), a construction tool that middle school students use for data analysis was the technological environment. In both environments, types of critical, creative, and statistical thinking were characterized. Two research questions were addressed: (i) How does the critical and creative thinking of middle school students using the software program Tinkerplots differ from their thinking while using a traditional paper and pencil format? (ii) How does the statistical thinking of middle school students using the software program Tinkerplots differ from their statistical thinking while using a traditional paper and pencil format? In order to answer the research questions students in grades 6 and 7 engaged in two tasks; one worked using the Tinkerplots software and the other worked using paper and pencil. This study was guided by the interplay of two theoretical perspectives. First, was the use of the

technology acting as an intellectual partner with the user, lending itself to the support of cognitive processes (Salomon, 1990). Second, was the concept of a computer functioning as a Mindtool, that is, a computer application that engages students in critical thinking (Jonassen, 2000). The findings of this study indicate that Tinkerplots appears to be a valuable asset in enhancing student's statistical thinking.

EXAMINING MIDDLE SCHOOL STUDENTS' STATISTICAL THINKING  
WHILE WORKING IN A TECHNOLOGICAL ENVIRONMENT

MELISSA ARNOLD SCRANTON

A Dissertation Submitted in Partial  
Fulfillment of the Requirements  
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

ILLINOIS STATE UNIVERSITY

2013

© 2013 Melissa Arnold Scranton

EXAMINING MIDDLE SCHOOL STUDENTS' STATISTICAL THINKING  
WHILE WORKING IN A TECHNOLOGICAL ENVIRONMENT

MELISSA ARNOLD SCRANTON

COMMITTEE MEMBERS:

Cynthia W. Langrall, Chair

Patricia H. Klass

Tami S. Martin

Edward S. Mooney

## ACKNOWLEDGEMENTS

I would like to thank Dr. Langrall and all my committee members. Dr. Langrall's guidance and support was instrumental to this project. She was there every step of the way, and I truly appreciate the time and effort she spent working with me. I am grateful to have the opportunity to work with Dr. Mooney, Dr. Martin, and Dr. Klass. A special thanks goes out to the Illinois State Math Department and Dr. Seelinger. I enjoyed my time living and working in Normal and have fond memories of my time as a graduate assistant.

I would like to thank my husband Jeff who stood by during this process. Without his support I could not have completed this task. I would also like to mention my two wonder children, Alexander and Amelia, who have so patiently waited for me to be able to play with again.

M.A.S.

## CONTENTS

	Page
ACKNOWLEDGEMENTS	i
CONTENTS	ii
TABLES	vii
FIGURES	viii
CHAPTER	
I. THE PROBLEM AND ITS BACKGROUND	1
The Software Program Tinkerplots	6
Purpose of the Study	9
Theoretical Perspectives	10
Computers as Intellectual Partners	10
Computers as Mindtools	12
Significance	13
Summary	14
II. LITERATURE REVIEW	15
Statistical Thinking and Learning	15
Comparing Distributions	16
Measures of Central Tendency	16
Variability	18
Data as a Distribution	22
Comparing Groups of Differing Sizes	23
Technology and Thinking Skills	25
Technology and Statistical Reasoning	30



III. RESEARCH DESIGN AND METHODOLOGY	41
Research Design	41
Participants	42
Pilot Study	42
Instrumentation	43
Tasks	43
Research Procedure and Data Sources	47
Software Training Sessions	47
Problem-solving Sessions	48
Paper and pencil task.	49
TinkerPlots task	49
Data Analysis	50
Data Analysis Overview	50
Data Analysis Phase 1	52
Critical and creative thinking	52
Data coding cycle 1: critical and creative thinking	53
Data coding cycle 2: critical and creative thinking	58
Data coding cycle 3: critical and creative thinking	59
Data coding cycle 4: critical and creative thinking	59
Data coding cycle 5: critical and creative thinking	60
Statistical thinking	60
Data coding cycle 1: statistical thinking	60
Data coding cycle 2: statistical thinking	61
Data coding cycle 3: statistical thinking	63
Data Analysis Phase 2	64
Data Analysis Phase 3	65
Summary	66
IV. FINDINGS	67
Findings by Case	68

Aspects of Critical and Creative Thinking Common Across Pairs	68
Pair 6A	70
Summary of paper and pencil session	70
Characterization of Pair 6A paper and pencil session	75
Summary of Pair 6A TinkerPlots session	78
Characterization of Pair 6A TinkerPlots session	87
Differences in thinking when working in the two mediums	91
Pair 6B	94
Summary of paper and pencil session	94
Characterization of Pair 6B paper and pencil session	97
Summary of Pair 6B TinkerPlots session	100
Characterization of Pair 6B TinkerPlots session	105
Differences in thinking when working in the two mediums	109
Pair 6C	111
Summary of paper and pencil session	111
Characterization of Pair 6C paper and pencil session	113
Summary of Pair 6C TinkerPlots session	115
Characterization of Pair 6C TinkerPlots session	125
Differences in thinking when working in the two mediums	129
Pair 7A	132
Summary of paper and pencil session	132
Characterization of Pair 7A paper and pencil session	136
Summary of Pair 7A TinkerPlots session	138
Characterization of Pair 7A TinkerPlots session	144
Differences in thinking when working in the two mediums	147
Pair 7B	149
Summary of paper and pencil session	149
Characterization of Pair 7B paper and pencil session	152
Summary of Pair 7B Tinkerplots session	154
Characterization of Pair 7B TinkerPlots session	159
Differences in thinking when working in the two mediums	163
Pair 7C	165
Summary of paper and pencil session	165
Characterization of Pair 7C paper and pencil session	171

Summary of Pair 7C TinkerPlots session	174
Characterization of Pair 7C TinkerPlots session	180
Differences in thinking when working in the two mediums	184
Findings of Cross-case Analysis	187
Cross-case Analysis of the Paper and Pencil Task	187
Representation	187
Context	188
The analysis cycle	188
Difference in group size	188
Method of analysis	189
Cross-case Analysis of the Tinkerplots Task	189
Representation	189
Context	190
The analysis cycle	191
Method of analysis	191
Difference in group size	192
Summary of Findings	192
V. CONCLUSIONS	194
Overview of Study	194
Answers to Research Questions	196
Research Question One	196
Context	196
Relationship to related literature	197
The analysis cycle	199
Relationship to related literature	201
Research Question Two	202
Representation	202
Relationship to related literature	204
Method of analysis	205

Relationship to related literature	207
Difference in group size	209
Relationship to related literature	210
Implications and Significance of the Study	210
Implications of the Technology	215
Did Tinkerplots function as a Mindtool and become an intellectual partner with the students in this study?	212
Implications for Instruction	215
Limitations and Constraints	219
Recommendations for Future Research	220
Closing Thoughts	220
REFERENCES	222
APPENDIX A: Opening Screen of Tinkerplots	236
APPENDIX B: Statistical Task: Study Habits	238
APPENDIX C: Isomorphic Task: Drivers' Education	240
APPENDIX D: Definitions of Critical and Creative Thinking: Jonassen (2000)	241
APPENDIX E: "Variation-Talk": Articulating Meaning in Statistics: Makar and Confrey (2004)	247

## TABLES

Table	Page
1. Chronological List of Codes for Pair 6A Paper and Pencil Session	76
2. Chronological List of Codes for Pair 6A Tinkerplots Session	88
3. Chronological List of Codes for Pair 6B Paper and Pencil Session	97
4. Chronological List of Codes for Pair 6B Tinkerplots Session	105
5. Chronological List of Codes for Pair 6C Paper and Pencil Session	113
6. Chronological List of Codes for Pair 6C Tinkerplots Session	126
7. Chronological List of Codes for Pair 7A Paper and Pencil Session	136
8. Chronological List of Codes for Pair 7A Tinkerplots Session	144
9. Chronological List of Codes for Pair 7B Paper and Pencil Session	152
10. Chronological List of Codes for Pair 7B Tinkerplots Session	159
11. Chronological List of Codes for Pair 7C Paper and Pencil Session	171
12. Chronological List of Codes for Pair 7C Tinkerplots Session	181

## FIGURES

Figure	Page
1. Opening Screen of Tinkerplots	8
2. Classical Music Task	45
3. Drivers' Education Task	46
4. Charts for Analysis Phases 1 through 5	56
5. Elements of Statistical Thinking	64
6. Back-to-Back Stemplot	72
7. Side-by-Side Circle Graphs	80
8. Multiple Circle Graph	82
9. Bar Graph	83
10. Side-by-Side Bar Graph	86
11. Bar Graph	101
12. Side-by-Side Bar Graph	102
13. Stacked Multiple Circle Graph	103
14. High and Low Scores	112
15. Side-by-Side Stacked Dot Plot	116
16. Side-by-Side Circle Graphs	118
17. Side-by-side Vertical Dot Plot	120
18. Red and Green Circle Graph	122

19. Side-by-Side Circle Graphs	124
20. Side-by-Side Vertical Dot Plot with Bins	140
21. Side-by-Side Horizontal Bar Chart with Bins	142
22. Side-by-Side Vertical Dot Plot with Bins	143
23. Fractions for Students Scoring Over 50	150
24. Fractions for Students Scoring Above and Below 50	151
25. Fractions for Students Scoring Above and Below 50	151
26. Side-by-Side Vertical Dot Plot	156
27. Four Quadrant Table on Desktop	166
28. (Recreation) Four Quadrant Table on Paper	168
29. Modified Four Quadrant Table on Paper	169
30. Bar Graph	175
31. Five Column Dot Plot	177
32. Dot Plot	178

## CHAPTER I

### THE PROBLEM AND ITS BACKGROUND

Today, we find ourselves living in a world that is dependent on information and the tools that deliver it to us (Ben-Zvi, 2000). Information affects the decisions made in our own homes, businesses, as well as the government sector. The delivery of accurate and timely data is critical for the decision making processes of our elected officials as well as for those who run the largest corporations around the world. Statistics plays a vital role in these processes and can have a tremendous impact on those decisions. Often times, those in positions of power will disagree over the same data but when these data are presented to the public via the news media, it is up to the individual at home to interpret the statistics presented. Today, the public is inundated with statistical data from TV, radio, and the internet. Thus, statistical literacy has become an important topic in our schools, preparing students to make sense of the massive amount of data they see on a daily basis.

The field of statistics has been transformed by the use of modern computing technology. Two decades ago, Taylor (1991) described three fundamental features of the computer that define its tremendous potential: state resurrection, time compression, and graphical representation of information. With regard to state resurrection, Taylor pointed out that because we can quickly edit and revise data, computers facilitate the production of higher quality products. Time compression relates to how easily computers remove the



monotonous labor involved with complex, time consuming tasks. Finally, Taylor pointed to the power of the computer to present information in a graphical form, a feature that has the potential to enhance the user's ability to interpret data. For the theoretical statistician, computers continually serve to increase the efficiency of their work and furthermore, today's powerful computers have opened new avenues of exploration (Biehler, 1993; Moore, Cobb, Garfield, & Meeker, 1995).

Although modern computing technology has only been available to schools for roughly 20 years (Becker, Ravitz, & Wong, 1999; Ben-Zvi, 2000; Chance, Ben-Zvi, Garfield, & Medina, 2007), it is obvious that during this brief period of time, computer technology has drastically increased the range, sophistication, and complexity of statistical activities that are possible in the classroom (Hawkins, 1997). Today's computers are more powerful than ever and when these machines are coupled with the latest software, learners are well equipped to "experience and explore all aspects of the statistical process – from planning the sampling or experimental design, through data collection, database management, modeling and analysis, to interpreting and communicating findings" (Hawkins, p. 3).

Advances in technology have led to rich applications that liberate students from complex computations, freeing them instead to explore, analyze, and think about statistical data (Chance et al., 2007). These tools are capable of displaying and manipulating statistical data in such novel and creative ways that they allow the student to focus on understanding of statistical concepts. Technology offers teachers new avenues and techniques for teaching statistics (Garfield & Burrell, 1997), which includes the use of computers, software, and the Internet. According to Friel (2008) these "are

essential tools for the practice in this domain” (p. 280).

The incorporation of technology into education has been endorsed by national organizations and encouraged by state and federal policies (Science and Engineering Indicators, 2010). To improve the performance of American students in science, technology, engineering, and mathematics, President Obama announced the “Educate to Innovate” program in November, 2009 (National Education Technology Plan, 2010). Technology is the central component of this plan as it will be the medium that provides powerful learning experiences, rich and engaging resources, and solid instructional content. For the study of mathematics, this plan recommends that “professional-level interactive graphing and statistical programs will make complex topics more accessible to all learners” (p. 14).

In response to the changes in society and technology, it is evident that both statistical thinking and instructional technology are important to the future success of mathematics education. With regard to teaching statistics, the National Council of Teachers of Mathematics (NCTM, 1989, 2000) has called for increased attention to be placed on statistical education in elementary, middle, and secondary schools, stating that “collecting, organizing, describing, displaying, and interpreting data, as well as making decisions and predictions on the basis of that information, are skills that are increasingly important in a society based on technology and communication” (p. 54). Furthermore, the NCTM has advised that a variety of technological tools should be used to perform routine calculations. These technologies can play a significant role in building conceptual understanding because they can provide divergent models of corresponding representations.

Recently, other major reports have outlined the importance of both statistics and computer competence in the education of our children and healthy maintenance of our workforce (Achieve, Inc. , 2004; College Board, 2006; Common Core State Standards, 2010; Guidelines for Assessment and Instruction in Statistics Education, 2005; NAEP, 2005). The consensus among these reports is that emphasis be placed on the process of statistical investigation, on using and constructing data representations, on data analysis, and the incorporation of appropriate technology to accomplish these goals.

Although the use of technology to teach statistics has been widely discussed and written about, there are relatively few empirical studies on this topic (Chance, Ben-Zvi, Garfield & Medina, 2007). Friel (2008) stated that “Much literature currently available with respect to statistics education is not research-based” (p. 291). Most of what is reported is based on observation and experience, explorations, or software features. Although there is a limited amount of empirical research that has explored how technological tools can be used to help students understand statistics, the studies that do exist illustrate the positive impact that technology can have on students’ cognitive development in learning statistics (Ben-Zvi, 2000; Hancock, Kaput, & Goldsmith, 1992; McClain & Cobb, 2001; Hammerman & Rubin, 2004). For example, Hancock et al. (1992) found that using the software program Tabletop supported middle school students’ understanding of a variety of graphs and representations. Ben-Zvi (2000) discovered that 13-year-old students' concept of time plots and trends in data were enhanced when they were exposed to spreadsheets. The author attributed this enhanced statistical understanding to several computer attributes such as the ability to plot, calculate, sort quickly and accurately, link multiple representations, simplify procedures, provide

feedback to the user, and transform representations into objects that students can easily manipulate. McClain and Cobb (2001) investigated how middle school students reason about data when using computer-based minitools. By the end of the study, over half of the students were able to describe a distribution as a proportion or percentage of the whole, and showed awareness of group propensities. Hammerman and Rubin (2004) observed that when using the software program TinkerPlots, middle and secondary schools students were able to view all of the data in a graphical representation. This affected the students' view of the mean and they had difficulty accepting it as being representative of the entire data set. Instead, they found that variability was a more compelling feature of the data and students were able to describe variability using the binning capability of TinkerPlots.

Jointly, these studies support the argument that computer tools can enhance statistical thinking. However, although computer based technologies have become widely available and are an important part of modern day curricula, it is not clear that these new technologies are being used effectively in the classroom or to their fullest potential (Lemke, Quinn, Zucker, & Cahill, 1998; McGraw, Ross, Blair, Hambrick, & Bradley, 2000; Technology Counts '99, 1999; Science and Engineering Indicators, 2010; U.S. Department of Education, The Final Report of the National Mathematics Advisory Panel, 2008). In fact, there is some evidence that technology is only being used superficially to teach statistics. Often, lessons consist of learning simple plots and finding standard measures of center, not on the creation of meaningful representations that aid in understanding and the decision-making process (Burrill, 1997).

Decision makers in schools must have information about the cost of technology to

be able to implement technology programs (Kelter & Ross, 1995; Redhead, 2001). Schools must consider not only the initial capital investment of software and hardware but also maintenance and support costs. The overall cost of such technology programs is more expensive than originally thought (Moses, 2002). According to Fickes (2004), “the cost to implement technology into our schools is enormous, tens of thousands per school, and millions per school system” (para. 2).

The increase in educational technology use in schools, coupled with evidence that technology is not being used to its fullest potential to teach statistics are disturbing findings leaving one to wonder about the learning return on our educational technology investment. Presently however, a formula that educators and policymakers could use to determine if the return on investment is suitable does not exist (Ringstaff & Kelley, 2002). Although there is some research in this area, we do not have a good grasp on the impact of technology on student’s thinking in a technological environment (Friel, 2008). The question of how valuable computer technologies are to the learning of statistics needs to be examined more closely. With a better understanding of the value of using computer technology teachers might be more inclined to use technology less superficially and in more meaningful ways. In addition, school systems will be able to make more informed decisions about technology purchases.

### **The Software Program TinkerPlots**

Today technological tools exist that can be used to teach statistics. TinkerPlots: Dynamic Data Exploration (Konold & Miller, 2005) is a promising, innovative computer program for the teaching and learning of statistics that middle schools students can use for data analysis (Bakker, 2002; Konold, 2002a, 2002b; Konold & Miller, 2001). Because

students can use this tool to build their own graphs, the authors of TinkerPlots refer to it as a construction tool.

Students who have no knowledge of formal statistics are able to begin to explore the world of statistics using TinkerPlots (Konold, 2002a). TinkerPlots allows students to create their own graphs and though they may not always resemble a typical bar graph or a scatterplot, students generate and manipulate these representations with ease. Although students may not know the difference between data types such as nominal, ordinal, or ratio, or have any knowledge of plot types or conventional statistical graphs, they are encouraged to explore data and answer questions with TinkerPlots. In this way, student understandings of statistical ideas and graphical representations gradually build from informal understandings to more formal concepts and standard representations.

When users open a dataset in TinkerPlots they are greeted with a screen that displays a random arrangement of icons (Figure 1 and Appendix A) (Bakker, 2002; Konold, 2002a, 2002b; Konold & Miller, 2001). Whether they are in the shape of circles or squares, each of the icons represents every case in the dataset. Users of TinkerPlots rely on a small set of buttons on a menu to stack, separate, and order the icons in a space free of axes. Without being forced to open a new tool, users can easily change the way the dataset appears on the screen to construct a wide variety of graphical representations.

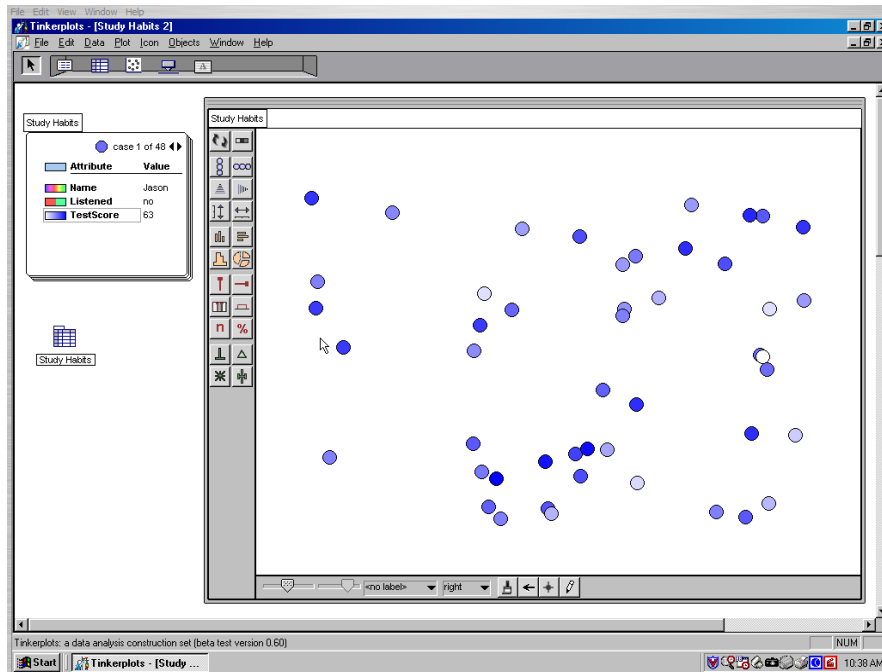


Figure 1. Opening Screen of TinkerPlots

What separates TinkerPlots from other more common software tools is that it does not teach context-bound knowledge, or provide mastery exercises (Bakker, 2002). Using the intuitive features of TinkerPlots, learners can manipulate a graph by rearranging data into different representations without having to have prior knowledge of conventional graphs. Instead, students can create bar graphs, pie charts, and scatterplots with relative ease. Most other statistical packages provide only standard statistical graphing functions that require learners to have some existing knowledge of statistics.

The pedagogy just described in TinkerPlots uses a *bottom-up* approach that is focused on how statistical reasoning develops in young students. Typically, other tools are based on the statistical knowledge of experts and this knowledge is presented to learners in its final form in a *top-down* manner. TinkerPlots provides much more than a simple interface where users go through the motions of printing text on paper. It is a

robust software tool that was designed to serve as a cognitive learning tool.

### **Purpose of the Study**

The purpose of this study was to gain a better understanding of how students think in a technological environment. This was accomplished by exploring the thinking of students while they worked in a technological environment and comparing this to the thinking they exhibited in a paper and pencil environment. In both environments types of higher order thinking and statistical thinking were characterized.

Two aspects of student thinking were examined. The first addressed the critical and creative thinking of students in two differing environments, technological and nontechnological. The conceptions of critical and creative thinking were defined in terms of Jonassen's (2000) model of critical thinking: "Critical thinking involves the dynamic reorganization of knowledge in meaningful and usable ways" and "creative thinking requires going beyond accepted knowledge to generate new knowledge" (pp. 27-28).

The software program TinkerPlots (Konold & Miller, 2005) was the technological environment and a traditional paper and pencil format was the nontechnological environment used in this study. The second aspect focused on the statistical thinking of students in these two environments. The definition of statistical thinking that guided this study "concerns the way one acts and what one thinks about during the course of statistical investigation" (Wild & Pfannkuch, 1999, p. 225). In particular, in my study I considered the way that students act and think while they engaged in a task that involved the comparison of two data sets. Both aspects focused on how students interpret and analyze data, and how these appeared to be affected by the two mediums. Two research questions were addressed:



1. How does the critical and creative thinking of middle school students using the software program TinkerPlots differ from their thinking while using a traditional paper and pencil format?
2. How does the statistical thinking of middle school students using the software program TinkerPlots differ from their statistical thinking while using a traditional paper and pencil format?

### **Theoretical Perspectives**

This study involved the interplay of two theoretical perspectives. First, was the use of the technology acting as intellectual partner with the user, lending itself to the support of cognitive processes. Second, was the concept of a computer functioning as a Mindtool, that is, a computer application that engages students in critical thinking. This theoretical perspective supports technology acting as an intellectual partner because Mindtools are computer-based learning environments that are designed to provide such partnerships (Jonassen, 2000).

The foundation on which both of these theoretical perspectives are based is constructivism. The theory of constructivism is focused on the process of how learners construct knowledge as they attempt to make sense of the situations in which they find themselves (Jonassen, 2000). How learners construct knowledge is the essence of constructivism and how learners build knowledge structures depends on the knowledge learners already possess. The creation of new knowledge structures is based on new experiences with objects and events that learners come in contact with.

### **Computers as Intellectual Partners**

Computers and software packages used in this study are not being utilized as

learning goals in and of themselves. More importantly, the use of computers is intended to help students to make sense of data. Undoubtedly, computers can have a tremendous impact in statistics classrooms with regard to the amount of time students spend on statistical calculations and graph construction. Computers can help to free our minds to concentrate on the actual problem at hand instead of being focused on the tedium of calculation. The true power of computers though is that they have the potential to enhance human intellectual activity in the classroom.

In this study I explored ways in which human intellectual performance was affected by the use of technology, focusing on the use of computers as cognitive technology tools. One definition of cognitive technologies states they are "any medium that helps transcend the limitations of the mind, such as memory, in activities of thinking, learning, and problem solving" (Pea, 1985, p. 168). The emphasis of my exploration was on whether cognitive technologies aided student thinking.

An important cognitive effect that technology can provide is defined by Salomon (1990) as the effect *with* a technology. It is possible that the use of computers opens doors for students to do things that they may never have had the opportunity to do before. Here, we are not so much concerned with the effects of computer technology on children as we are with what computers enable students to actually *do*, or what is referred to as the aspect of effects *with* technology. Computers afford students opportunities to manipulate functions, test any plausible hypothesis, and to investigate anything they can invent in their imaginations. In this way, computer programs offer students "an intellectual partnership that transcends the limitations of human cognition" (Pea, 1985). What's more, computers do not just serve to strengthen the abilities of children but they also

teach them new ways of looking at tasks.

### **Computers as Mindtools**

Although Salomon (1990) stated that computer based technology has the potential to support and extend human intellectual performance, Jonassen (2000) took this view a step further. He referred to Mindtools as computer-based, constructivist learning systems designed to provide an intellectual partnership with the students who use them (Jonassen, 2000). Mindtools promote critical thinking and higher order learning by using tools such as databases, networks, spreadsheet software, system modeling tools, World Wide Web search engines, and multimedia software, to name a few.

Using computers as Mindtools allows students to explore the world and learn to think in ways that otherwise they could not (Jonassen, 2000). Mindtools are in essence, knowledge construction tools where the student learns *with*, not from the Mindtool in use. In the environments that students are exposed to by these tools they find themselves actively interpreting the outside world in new ways, while at the same time reflecting on these experiences. When computers function as Mindtools, the thinking processes of the student are supported, nourished, and extended by these technologies. The use of Mindtools engages students when asked to represent what they know but it goes much deeper than this. When students are fully engaged in complex critical, creative and problem-solving situations, thinking skills can be significantly extended by the use of Mindtools.

Mindtools are not designed to reduce the amount of mental processing that a student will need to do while using them (Jonassen, 2000). Instead they are designed to more effectively use the mental efforts of students. When students begin to use Mindtools

they quickly understand that they must have a good understanding of the subject matter and the problem domain in which they are working for Mindtools to have the greatest effect. It is only when students make the conscious decision to work in concert with these tools, with the understanding that they will be required to think deeply in the process, that these tools will be of the greatest benefit.

Utilizing a constructivist approach, Mindtools help learners to represent, manipulate, and reflect on what they know with computers or other technologies, environments, and activities (Jonassen, 2000). In the environments that Mindtools provide, students actively engaged in creating knowledge for themselves that reflects how well they comprehend and understand information rather than relying on the representations of information by others. When students proceed to build knowledge that is based on their beliefs and interpretations of the world around them, the possession of this knowledge is uniquely their own and it is more firmly held onto over time.

From a theoretical perspective, Mindtools provide a defining point of view. The technology in the context of this study was thought of as a Mindtool. Instead of simply thinking of TinkerPlots as a construction tool, it was thought of as a Mindtool, which is more specific. The primary focus of Mindtools is to develop meaningful thinking through the use of computers and they are designed to provide an intellectual partnership with the students who use them. Participants in this study were encouraged to form an intellectual partnership with the computer, which could afford them new ways of looking at data and open doors to do things that they may never have had the opportunity to do before.

### **Significance**

The purpose of this study was to gain an in-depth understanding of how students

think in a technological environment. The intent was to characterize the types of higher order thinking and statistical thinking that students exhibit in technological and nontechnological environments. The relationships documented between working in a computer environment and paper and pencil environment will help to establish the benefit of using computers to teach and learn statistics.

Student thinking in a technological environment is an emerging field of study and there are limitations to the existing research. This study will help to fill the gaps in the research on statistics and technology. To address the limitations of the current research, this study utilized the most up-to-date computer software package to teach statistics that was developed with the intent of being used with constructivist mode of learning. It will also be one of only a few studies where computer technologies are the central component.

### **Summary**

In this study, students were given the opportunity to explore an open-ended technological environment and create their own knowledge structures, based on their own experiences. The theoretical perspectives utilized a constructivist approach by providing an environment where learners were actively engaged in creating knowledge for themselves that reflected how well they comprehend and understand information rather than relying on the representations of information by others.

## CHAPTER II

### LITERATURE REVIEW

The literature review appropriate for this study will include two main components: (1) research on statistical understanding, and (2) research on the use of computer technology. The following review outlines major research pertaining to critical issues in statistical thinking and learning, the use of computer technology in the development of thinking skills, and the use of computer technology in the development of statistical reasoning at the middle school level.

#### **Statistical Thinking and Learning**

For this study, middle school students were engaged in tasks that involved comparing distributions. There is now a growing body of literature that focuses on middle schools students' understanding of comparing distributions (Ben-Zvi, 2004; Cobb, 1999; Gal, Rothschild & Wagner, 1989; Hammerman & Rubin, 2004; Konold & Pollatsek; Konold et al., 2002; Makar, 2004; Makar & Confrey, 2002a, 2002b, 2004, 2005; Petrosino, Lehrer, & Schable, 2003; Shaughnessy, Ciancetta, Best & Canada, 2004, Watson & Moritz, 1999). This section of the review includes concepts related to comparing distributions. There are several areas in research that are pertinent the understanding of comparing distributions, namely measures of central tendency, measures of variation, and distribution. This chapter consists of the subsections related to these areas. Only studies most closely related to middle school students' understanding of

comparing distributions were reviewed in order to help stage the current study.

### **Comparing Distributions**

One of the most basic tasks in statistics is to examine differences in two sets of data. Comparing distributions requires students to evaluate and connect factors such as central tendency and variation within representations. Konold and Higgins (2003) describe comparing distributions as “heart of statistics” (p. 206). Comparing distributions supports the exploration of centers, variability, and distribution (Bakker & Frederickson, 2005; Cobb, 1999; Konold & Higgins, 2003; Konold & Pollatsek, 2002; Petrosino, Lehrer, & Schauble, 2003; Watson & Moritz, 1999).

### **Measures of Central Tendency**

One of the first topics examined by researchers about students’ statistical thinking was the average or mean. These measures are essential to being able to compare differences between data sets. Many research studies examine how students think about measures of central tendency, specifically the average (mean), mode, and median (Mokros & Russell, 1995; Watson & Moritz, 2000; Strauss & Bichler, 1988). These findings of these studies suggest that students struggle with the concept of the mean of a distribution. Research has shown that students do not use means to compare data sets, even when familiar with them (Bakker, 2004; Bright & Friel, 1998; Gal, Rothschild, & Wagner, 1989; Konold & Pollatsek, 2002; Rubin & Hammerman, 2006).

Gal, Rothschild, and Wagner (1989) conducted a study to explore elementary students’ intuitions and statistical reasoning strategies when comparing two sets of data. The participants were 62 third and sixth graders. The context of one data set was the distance that frogs jumped in a frog jumping contest. The other data set was student

scores on a school exam. Participants were asked to determine if the groups in the two different contexts did about the same as one another or if one group did better. The characteristics of the distribution were manipulated so pairs of participants received variations of the same distributions. For example, one comparison involved two data sets with one shifted far to the right with no overlapping values. Another pair of participants received distributions with one skewed left and the other was skewed right, creating a difference in the locations of the centers. The researchers observed that the participants used an array of strategies to compare the two data sets. However, few students reasoned statistically in their analysis of the data. The students failed to use the concept of average when comparing the two data sets. Even the sixth graders familiar with the concept of average failed to apply it to their comparisons.

Watson and Moritz (1999) conducted a study to explore student understanding of comparing two data sets. Participants were 88 students in third through ninth grade. The researchers found that not many of the students were able to reason about the mean. They thought that prior student experiences with calculating the mean using the standard procedure may be interfering with their ability to apply the concept. The researchers found that comparing two distributions was motivating and suggested it be included in the curriculum starting in third grade using same size data sets.

Konold and Pollatsek (2002) recognized that much prior research had focused on the average; however, this research did not connect different interpretations of the average. As a response to this, Konold and Pollatsek (2002) created a hierarchical four step framework of interpretations of averages: typical value, fair share, a way to reduce data, and as a “signal amid noise.” A typical value is the most frequent, usual, or central



data value. A fair share is where the data is equally distributed between groups or individuals. A data reduction is the calculated average where the data is reduced to a single data value. These three interpretations (typical value, fair share, data reduction) had all been discussed in the literature on this topic. Konold and Pollatsek added a new interpretation “a signal in noise.” The stable component in the data that one is trying to identify is the signal. The variability around the stable component is the noise. The researchers proposed that the “signal amid noise” was the most meaningful for comparing two data sets and suggested that elementary students be introduced to the average when comparing data sets.

### **Variability**

Variability is considered the hallmark of statistical thinking (Wild & Pfannkuch, 1999). Variability is defined as “the propensity for something to change, whereas variation is a measurement of that change” (Shaughnessy, 2007, p. 972). Some researchers differentiate the terms variation and variability (Reading & Shaughnessy, 2004). However, others make no distinction between the two terms and use them interchangeably (Makar & Confrey, 2004b). Research in students understanding and development of variability is just in beginning stages (Watson et al., 2003). Research on variability has focused on variation in data, conceptions of variability in samples, and variability in the comparison of distributions (Shaughnessy, 2007). The focus of this section will be on using variability to support the comparison of distributions.

Ben-Zvi (2004) explored how two seventh-grade student reasoned about variability when comparing distributions of equal size. The participants were compared the lengths of 35 Israeli surnames to the length of 35 American surnames using Excel.

Seven stages that students progress through were documented that described variability within and between groups; Stage 1 On what to focus: Beginning from irrelevant and local information; Stage 2 How to describe variability in raw data; Stage 3 How to formulate a statistical hypothesis that accounts for variability; Stage 4 How to account for variability when comparing groups using frequency tables; Stage 5 How to use center and spread measures to compare groups; Stage 6 How to model variability informally through handling outlying values; Stage 7 How to notice and distinguish the variability within and between the distributions in a graph (p. 48). Ben-Zvi found that understanding of variability in the comparison of the two groups was aided by students' ability to view a distribution as a global entity. This view of a distribution seemed to be a prerequisite for describing two distributions that had similar shape and variability, however shifted to the left of right of each other.

Shaughnessy, Ciancetta, Best, and Canada (2004) conducted a study on focusing on variability when comparing distributions. Participants were eight were middle school students and 16 were high school students. Responses for one of the tasks, Movie Wait-Time, were coded into six categories: specific data points, variation, centers, distribution, informal inferences, and context. Responses could fall into multiple categories. Results showed that most students attended to both centers and variation. About two-thirds of the students said the data sets were different despite having the same mean and median. Most provided reasons that could be categorized in multiple ways with a majority linked to variation. About one-third of the students included personal contexts and past experiences in their responses.

Konold et al., (2002) conducted a study to explore how middle school students

reason when comparing data sets. Participants were students in seventh and eighth grade science classes. The students examined the number and type of animals killed on local roads, and developed questions and hypothesis, and investigated their questions by analyzing subsets of the data. The students incorporated the idea of what Konold et al., (2002) deemed 'modal clumps' into their analysis. These bumps were the middle hump of mound-shaped distributions. Students often used the mean as a way of characterizing a modal clump due to its location in the middle of the set. The modal clumps were generally ranges in the middle of the data sets that included the mode and had a higher percentage of data than either of the other two partitions. Konold et al., referred to the modal clump as a "range of values in the heart of a distribution" (p. 1). Konold and his colleagues suggest that modal clumps are a good starting point for communication about the typicality and variability within and between data sets. Evidence was presented that suggest that instruction on summarizing and comparing data should begin with the identification and discussion of modal clumps as opposed to measures of central tendency.

Another critical issue in the process of statistical analysis is the concept of group propensity, that is, the "intensity or rate of occurrence of some characteristics within a group composed of elements that vary on that characteristic" (Konold, Pollatsek, Well, & Gagnon, 1997, p. 2). The majority of students could not compare groups in terms of means or relative frequencies when conducting data analysis and they often failed to reason multiplicatively. Konold et al. asserted that the ability to identify propensities in the data is central to a statistical perspective.

Makar and Confrey have conducted several studies on how teachers' compare

data sets (Makar, 2004; Makar & Confrey, 2002a, 2002b, 2004, 2005). In Makar's 2004 study, preservice secondary mathematics and science teachers used informal non-standard language to describe a variation while comparing data sets. Teachers used expressions such as clustered, clumped, bunched up, scattered, evenly distributed, dispersed, spread out, grouped, and gathered to describe distributions. The researchers believed that the informality of the teachers' internalization of these concepts may not be able to support their understandings of more complex topics. They found that even if teachers struggle with comparisons between distributions even when they use formal measures.

In a 2005 study by Makar and Confrey, the researchers referred to the preservice teachers used of informal non-standard language when describing distributions as "variation talk." Terms the teachers used to describing center, spread and variation included "bulk of this data," "scattered," and "bunched." The findings suggest that the informal language is an important form of communications. Informal language and intuitive ideas may develop prior to formal definitions, and should be encouraged.

Cobb (1999) analyzed 29 seventh grade students' mathematical reasoning during a ten-week teaching experiment designed around the overarching idea of distribution. Over the course of 34 lessons the students made decisions or judgments based on their descriptions and comparisons of data sets. The initial lessons required the students to describe single data sets and later the students compared two or more sets of data. Both types of activities required the use of two computer-based mini-tools. The tools allowed the students to organize and partition the data.

In the initial lessons, some students answered every question with the mean. As

the study progressed, students started exploring qualitative characteristics of the data. These characteristics helped the student's begin to formulate conceptions about distributions, and they began to look beyond individual data points. However, the students were only looking at numeric frequencies, rather than relative frequencies, which resulted in their continuation of additive reasoning.

The second set of lessons focused on reasoning about the data multiplicatively. Multiplicative reasoning occurs when students partitioned the data into pieces. Then students use proportional reasoning when using a proportion (or percentage) of individual data points in part-whole terms. At this point in the study the students began to reason multiplicatively and discuss proportions within the data sets. The students also described global patterns in the data using phrases such as "hill shaped" to describe a data. This signified a shift in focus from single data points to a more global view of the data.

When using the first mini-tool, students used predominantly additive reasoning, and transitioned to multiplicative reasoning using the second mini-tool. This was accompanied by a shift in the class discussions from practical decisions or judgments to strategies on how to organize the data that supports the decisions and judgments being made.

### **Data as a Distribution**

In statistics courses, one of the first and most important topics introduced is distribution. Research has shown that the concept of distribution plays a major role in understanding how data behaves in graphical situations and is connected to the reasoning and understanding of variability (delMas & Liu, 2005; Makar & Confrey, 2003; Reading & Reid, 2004). The importance of distributions in understanding statistics has been well

articulated by researchers (Bakker and Gravemeijer, 2004; Ben-Zvi & Garfield, 2004; Konold & Higgins, 2003) and is often thought of as a unifying concept in statistics (Bakker, 2004; Cobb, 1999; McClain, Cobb, & Gravemeijer, 2000).

There is a growing body of evidence that in student's initial experiences with data they "tend to focus on describing individual data points, or clusters of similar individuals." (Konold & Higgins, 2003, p. 202). An important step in considering data from a statistical perspective is to make a shift from thinking about the data in local ways to considering the data as a whole entity. The foundation to being able to view data globally is considering a data set as a whole entity.

Ben-Zvi and Arcavi's (2001) research indicated that viewing data locally often constrained students from analyzing the data globally. However, there were times when viewing data locally was foundational for their global assessment of the data. The researchers concluded that it was important that students need to be able to move back-and-forth between the two.

According to Mokros and Russell (1995) "Until a data set can be thought of as a unit, not simply as a series of values, it cannot be described and summarized as something that is more than the sum of its parts" (p. 35). The ability to view data as a unit is necessary for an understanding of distribution, which in turn in needs to be able to reason about variation. When able to view a set of data as a unit, the data can be described by trends, patterns, shape, center and spread. The identification of these characteristics are used to make comparisons between data sets.

### **Comparing Groups of Differing Sizes**

Several studies on students' comparing data sets of unequal size have been

conducted. Watson and Moritz (1999) interviewed 88 students from Tasmania and South Australia in grades three through nine, to examine student thinking when comparing graphs. The majority of students used more than one solution strategy, and were more successful than the students who only used one strategy. Only a few students found the mean using the standard algorithm. The researchers believe that students should have the opportunity to compare data sets intuitively based on a distributional perspective. This will help students from blindly applying formula without their understanding of them.

Petrosino, Lehrer, and Schable (2003) conducted an assessment tasks on 14 students involved in comparing distributions of unequal size. The students were asked to decide which of two basketball players should be selected for an All-Star game based the number of points they scored in several games. Over half of the students chose the player who scored the most points. After probing by the interviewer, half of these students switched their response to the player who played in the most games. In addition, two students chose the students with the largest range of scores, reasoning that he had more scores.

Petrosino, Lehrer, and Schable (2003) conducted another assessment with the same 14 students comparing two data sets that had different shapes and equal ranges. The students were asked to decide which of two types of soil (light and dark) resulted in better tree growth. The students were given height data on trees grown in both types of soil. Six students determined the soil type did not make a difference, four students calculated that averages (which were the same). And the remaining students reasoned about similar centers based on the variation in the data sets. The student used other strategies including analyzing and comparing the extreme scores, comparing medians, comparing the

percentage of trees that are average height or taller and comparing average spreads.

In another study, Watson and Mortiz (1999) examined how students compared two data sets by presenting students with pairs of distributions representing math test scores from two different classes. The students were asked to determine which of the classes did better, or if they scored about the same. Many students used a visual approach which involved looking for the tallest bar in the representations and looking for differences between graphs. A common numeric strategy employed by students was to add all the scores for a class together. This strategy was not helpful for working with data sets of unequal size. The researchers found that the students who could successfully compare unequal size groups were multiplicative reasoners. Students who were not multiplicative reasoners struggled with groups of unequal size. These students were much more likely to employ additive reasoning as a strategy, comparing numbers of students as opposed to the frequency of students. In light of these findings, multiplicative reasoning has been established as an important factor in the comparison of data sets of unequal size.

### **Technology and Thinking Skills**

Many research studies have examined the impact of tutoring systems technology, such as computer-based instruction (CBI), computer-aided instruction (CAI), and drill-and-practice software on student achievement. These studies have shown that when students from preschool to higher education use technology, it can improve their basic skills and test scores on achievement tests. (Almeqdadi, 2005; Ash, 2005; Barrow, Markma & Rouse, 2008; Bayraktar, 2002; Clements & Sarama, 2009; Institute of Education Sciences, 2009; Kulik, 1994; Kulik & Kulik, 1986; Lowe, 2001/2002; Mann, Shakeshaft, Becker, & Kottkamp, 1999; Pearson Digital Learning,



2006; Ringstaff & Kelley, 2002; Roschelle et al., 2010; Schacter, 1999; Sivin-Kachala, Bialo, & Langford, 1998; Timmerman & Kruepke, 2006; Williams, 1996).

One study that typifies this impact was a meta-analysis of over 500 studies of computer tutoring systems conducted by Kulik in 1994. Students who used CBI had percentile improvements between 9 and 22 percent on achievement tests compared to control groups. Typically, these students scored in the 64<sup>th</sup> percentile while those who did not use computer-based instruction scored in the 50<sup>th</sup> percentile.

Another meta-analysis was conducted by Hsu (2003) to gauge the effectiveness of computer-assisted instruction (CAI) when used to teach college level introductory statistics courses. Included in the study were 25 primary studies from journal articles, ERIC documents, and dissertations. The meta-analysis produced an overall effect size estimated to be 0.43 which indicates that CAI can have a positive effect when used to teach introductory statistics courses.

Personal computers clearly have great potential as classroom tools (Moore, 1997). While most of the studies regarding computer aided instruction have examined student achievement using an outcome measure of course grade or performance on standardized achievement tests, there is now a growing base of research on how these technologies may be used with students during instruction. Over a quarter of a century ago, Pea (1987) said that technology tools can be used to enhance understanding. Many researchers have turned their attention to computer-based technology being used as a tool for problem solving, understanding of concepts, and critical thinking.

When computer-based technology is used in this manner, students use the information they have gathered to interpret, analyze and solve problems. However there

is one caveat; used in a constructivist mode, it is often difficult to measure the impact of technological instruction on student learning (Ringstaff & Kelley, 2002). The difficulty lies in the fact that many of the existing assessment tools are insufficient when it comes to accurately measuring the skills that computer-based technologies enhance such as critical thinking, higher-order thinking skills, and problem solving. Unfortunately, the many studies of technological tutoring systems offer little insight into the impact on students learning with this technology to develop thinking skills. Nonetheless, a limited number of studies exist that illustrate the positive impact that technology can have on higher-order critical thinking skills.

One such study in mathematics, conducted by Wenglinsky in 1998, examined the affects of simulation and higher order thinking technologies. Across the country, 6,227 fourth-grade students and 7,146 eighth-grade students performance was assessed on the National Assessment of Education Progress (NAEP) test. The eighth-grade students who used these technologies were found to perform as much as 15 weeks above grade level in math scores as measured by NAEP.

Cousins and Ross (1993) investigated the use of computer technology to improve students' correlational reasoning abilities. Four hundred eighty-three 9th- and 10th-grade geography students were asked to find a relationship between two continuous variables. After the treatment, students were measured on four skill dimensions: organizing, locating, synthesizing, and concluding. Four student treatment groups outperformed a control group on the four dimensions of correlational reasoning.

Hopson, Simms, and Knezek (2001/2002) investigated the effect of a technology-enriched classroom on fifth and sixth-grade students' development of higher-order

thinking skills, as defined by the analysis, synthesis, and evaluation levels of Bloom's Taxonomy. Results indicated that the treatment group exhibited a higher level of evaluation skills, but no significant difference on the analysis and synthesis subtests. The authors concluded that the effects of a technology-enriched classroom on students' attainment of higher-order thinking skills are minimal but positive.

Ben-Zvi & Friedlander (1997) conducted a three year project whose aim was to provide middle school students with more meaningful statistics learning, they discovered that learners' cognitive loads were shifted away from activities that involved drawing graphs and performing calculations to other activities that employed higher-order thinking skills while using computer tools. It is encouraging to note that learners' cognitive resources were redistributed to other areas such as software, statistical reasoning and methodology while using computer technology.

Polly (2011) describes how third grade students used an activity on the NCTM's Illuminations website to deepen their understanding of algebra and increase their higher-order thinking skills. The students used the "Pan Balance – Shapes" activity to investigate equality by placing shapes on the balance, which became level when the shapes on both sides had the same value. The author equated the processes the students used to the three highest levels on Bloom's revised taxonomy for thinking skills (Churches, 2008). The process of communicating about math with was equated to analyzing, determining the value of shapes was equated to evaluating, and writing equations was equated to creating. The author concluded that this activity deepened students' mathematical understanding as well as promoted higher-order thinking skills.

Einav and Nenad (2012) explored how high school students who used an

interactive tool called GeoGebra as well as critical thinking skills enriched their understanding of conditional probability and Bayes' theorem. The author's found that by manipulating objects in the dynamic learning environment and applying critical thinking skills, students were able to comprehend the concepts of Bayes' formula. These two elements, GeoGebra and critical thinking skills, enabled students to comprehend the association of base rates and conditional probability.

Shaharuddin, Tasar, and Shukor (2012) developed a web-based, interactive framework designed to boost critical thinking skills. To evaluate the effectiveness of the framework, the authors developed web-based simulations that were intended for learning educational communication and networking. The case study, which involved 21 higher education students, showed that the impact of the framework on the students' critical thinking skills was positive.

There is a great need to increase the development of higher-order critical thinking skills. The use of technology as a tool to promote student understanding of mathematics was recommended in the NCTM's Principles and Standards (2000). Bloom's Revised Taxonomy supports the use of technology in mathematics to advance student understanding and high-order thinking skills (Churches, 2008). The three highest levels of Bloom's Revised Taxonomy are creating, evaluating, and analyzing. The limited number of studies pertaining to the impact that technology used for mathematics learning can have on improving higher-order thinking skills suggests this is a gap in the existing literature.

Research from other fields of study exists that also demonstrates the impact that technology can have on higher-order thinking skills. One such study in science education,

examined whether teaching for the acquisition of higher-order thinking skills enhanced high school students' critical thinking (Barak, Ben-Chaim & Zoller, 2007). Students were divided into three groups; an experimental group of science students, a control group comprised of science majors, and another control group made up of non-science majors. The experimental group was taught using the strategies designed to enhance higher-order thinking and the two control groups were taught using traditional methods. The author's found that the first group showed a statistically significant improvement on critical thinking skills compared to the control groups.

Jointly, these studies strongly support the argument that computer tools can become an intellectual partner with the student. The unfortunate reality is that teachers are not using computers in the most effective ways. In Wenglinsky's (2005/2006) examination of the National Assessment of Educational Progress (NAEP) from 1996, 1998, and 2001 showed that using computers for solving complex problems produced greater benefits than using them for drill-and-practice. However, it was reported that only 30% of mathematics teachers used computers for tasks that involved higher-order thinking skills.

### **Technology and Statistical Reasoning**

As shown in the previous section, the integration of technological tools in the classroom can have a major impact on student learning. They can promote a "richer, powerful, and flexible learning environment in which students are active learners of statistics" (Ben-Zvi, 2000, p. 149). When computer technologies are focused on assisting students with organizing, analyzing, and displaying data, researchers believe that these technologies can help students learn mathematical skills such as graphing and statistics.

Research into the area of statistics has just begun and much more will need to be done to determine the appropriate use of technology in learning statistics (Shaughnessy, Garfield, & Greer, 1996).

Research on the use of technology to enhance statistical thinking and learning is complicated in that many studies now involve a technology as an integral component of the research, but not as the sole focus. Thus, as technology is intertwined into these studies, its impact is often difficult to discern. Although research in this area is limited, there are a handful of studies that highlight using technology in the teaching and learning of statistics.

Tabletop, a computer software tool for middle school students that can be used to build, explore, and analyze databases, was used by Hancock et al. (1992) to explore the challenges associated with authentic inquiry for fifth through eighth grade students. They discovered that Tabletop was able to support student understanding of a variety of graphs, making representations intelligible for students.

Ben-Zvi (2000) explored the effects of access to computer spreadsheets on 13 year old students' construction of meaning in statistics by having them produce, interpret, transform, and link data representations. The researcher discovered that students' understanding of time plots and trend in data were enhanced. Several attributes of computers such as the ability to quickly sort, plot, and calculate while providing immediate feedback help students transform representations into objects they can easily manipulate.

McClain, Cobb, and Gravemeijer (2000) looked at how middle school students came to reason about data in an exploratory data analysis environment. In their teaching

experiment, they used a computer minitool that assisted students in ordering, partitioning, and organizing sets of data. Through this process, they focused on comparing data sets through the use of multiplicative reasoning strategies rather than additive reasoning strategies. By the end of the study, over half of the students were able to describe a distribution as a proportion or percentage of the whole, and showed awareness of group propensities.

During the last ten years there has been an explosion of technology-based tools that learners can use to investigate statistics. The most notable of these are TinkerPlots (Key Curriculum Press, 2005), Fathom (Key Curriculum Press, 2005), and CPMP-Tools (Keller, 2006). These tools stand out because they allow both teachers and students to interact with data while performing statistical investigations in unconventional ways. TinkerPlots and Fathom are categorized as “landscape-type” tools (Baker, 2002) that allow learners the freedom to move between vast arrays of data representations in an environment where the learning objective is not assumed like it is in tools such as CPMP-Tools which are categorized as “route-type” tools (Baker, 2002). Route-type tools are designed for specific learning objectives that are part of a broader mathematics learning environment.

Research about how these learning environments affects students’ statistical thinking and reasoning has just begun. The software program TinkerPlots was the technological environment used in this study. TinkerPlots is a dynamic graphing tool that was created for middle school students that is based on constructivist learning theories (Konold, 2002, 2007). The rich learning environment that TinkerPlots provides builds on the statistical knowledge of those using the tool already have. It does this by allowing

them to explore groups of data in interesting ways. Data points are displayed as colored circles or squares where the shade of color has meaning as well as where the shape appears in relation to other data points. Using the intuitive features of TinkerPlots, learners can manipulate a graph by rearranging data into different representations without having to have prior knowledge of conventional graphs. Instead, students can create bar graphs, pie charts, and scatterplots with relative ease. When these activities are performed, learners are able to discover patterns in the data that allow them to reach conclusions about the data using the representations they created themselves.

Research that used TinkerPlots as a tool utilized a variety of statistical contexts to engage learners. These studies looked at how TinkerPlots served as a vehicle for examining the way in which the software assisted students to statistically analyze data. In addition, some studies focused on the development of teacher understanding of statistics. Rubin and Hammerman (2006) reported that results for teachers using TinkerPlots were similar to those for students, which makes the knowledge gained from studies involving teachers useful when considering student thinking and understanding.

Several studies using TinkerPlots have focused on gaining insights into how students develop informal inference (Ben-Zvi, 2006; Papanastasiou & Meletiou-Mavrotheris, 2008; Watson, 2008; Watson & Donne, 2008; Watson & Wright, 2008). Findings from these studies support the view that TinkerPlots can promote the development of informal inference. Two of these studies are summarized below.

Ben-Zvi (2006) conducted a study focusing on how students develop formal ideas of inference. Seventy-five primary school students' (ages 10-12) evolving ideas about statistical reasoning were investigated using TinkerPlots. Ben-Zvi found an advantage of



using TinkerPlots was that students were able to use it successfully as an argumentative tool for presenting their ideas, as opposed to just a representational tool. However, students often focused on individual data points, rather than treating the distribution as a whole, and seeing the aggregate view of the data.

Paparistodemou and Meletiou-Mavrotheris (2008) conducted a study to determine how students' ideas about informal inference can be developed by using TinkerPlots. Twenty-two third grade students in Cyprus made data-based arguments using TinkerPlots as an investigational tool and medium for social activity. Paparistodemou and Meletiou-Mavrotheris found that TinkerPlots can support the development of inferential reasoning in young children. They noted that most of the time the children tried to find relationships between the variables and draw conclusions. This occurred most often when students could relate to the activity and found it interesting. The researchers also found that being able to visualize the data helped children proportionally reason about the data and identify trends and group differences. Overall, the children's statistical conceptual understanding was enhanced due to "the ability to operate quickly and accurately, to dynamically link multiple representations, to provide immediate feedback, and to transform an entire representation into a manipulable object enhanced students' flexibility in using representations" (Paparistodemou & Meletiou-Mavrotheris, 2008, p. 101).

Several studies using TinkerPlots have focused on understanding and comparing data distributions (Bakker & Frederickson, 2005; Hammerman & Rubin, 2003; Hammerman & Rubin, 2004; McClain, 2002; Rubin & Hammerman, 2006; Rubin, Hammerman, & Konold, 2006). Findings from these studies highlight the statistical

analysis techniques used to compare two sets of data. Several of these studies are summarized below.

Hammerman and Rubin (2003) conducted a study to investigate how computer visualization tools can enhance learning. Middle and high school teachers used TinkerPlots and Fathom as the computer environment in this study. The researchers found several ways that people manipulated the data to make it more understandable. Teachers categorized the data into bins, used cut points within the distribution, identified gaps, and clumped similar values together. They also created pie graphs and used percents and ratios to compare groups with different sample sizes. These techniques equalized the difference in group size and allowed comparisons to be made. The researchers noted that the teachers created many different graphs quickly, analyzing them and then moving on to another graph.

Hammerman and Rubin (2004) conducted a study to investigate how teachers handle variability in data distributions with a goal of discovering how teachers use TinkerPlots to compare groups. Eleven middle and high school teachers participated in the study exploring how data visualization tools can enhance learning. Hammerman and Rubin found using measures of center was less common than other analysis techniques. They stated:

Using measures of center was by far less common than the other strategies we have described in this paper. Our experience is, in fact, that seeing a distribution makes it harder to accept a measure of center, especially a mean, as being representative of the entire distribution, with its particular spread and shape. In this sense, the binning capacity of TinkerPlots we believe, filled a “need” these

teachers had to describe a distribution when the variability was perhaps more compelling than any measure of center. (p. 36-37)

The researchers noticed that a need for multiplicative thought was created by using bins for the two groups of data of unequal size. Teachers had to compare a bin in one distribution with the same bin in another, often with a different number of elements in them. At the end of the study, most of the teachers were able to use multiplicative reasoning some of the time. The researchers found that the visual features of TinkerPlots supported the development of the teachers' understanding of variability and aided them in the creation of complex and sophisticated arguments.

Rubin and Hammerman (2006) conducted a study to investigate how the shape of data affects middle and high school teacher's comparison of two data sets using computer visualization software. Six of the teachers used TinkerPlots and three teachers used Fathom to analyze the two different sets of data. One set of data was symmetric about the height and gender of a group of Australian teenagers. The other set of data was skewed about the money earned from part-time jobs for this group of teenagers. Rubin and Hammerman found that teachers analyzed the two sets of data differently. For the symmetric distribution teachers used measures of central tendency and for the skewed distribution they used cut points to determine how much data was above and below the cut point.

McClain (2002) conducted a study to investigate how learners visualize authentic data sets using TinkerPlots. To compare two data distributions participants created graphs with bins, and then used proportional reasoning to analyze the connection between bin size and group size.

Several studies have compared student work using only paper and pencil to their work in TinkerPlots (Rubin, 2002; Meletiou-Mavrothersis, Papparistodemou & Stylianou, 2009; Watson & Donne, 2009). Findings from these studies highlight the differences in statistical analysis techniques that were used in each environment. These studies are summarized below.

Rubin (2002) conducted a study where teachers compared the lifespan of two types of batteries by completing the task in TinkerPlots, and also completing it by-hand with only paper and pencil. When the teachers worked the problem with paper and pencil, they calculated the mean lifespan for each brand and then compared the means for both groups. The teachers' focus during the paper-based session was on the algorithmic calculations. It was noted that they spent a lot of time on their calculations. However, the teachers were not satisfied with this measure due their observation of outliers. When using TinkerPlots, the teachers created a graph with the lifespan of each battery separated into six bins. The teachers observed there were four outliers outside the modal clump and suggested these should be deleted. Rubin suggested that TinkerPlots gave the teachers a different view of the data and that their conclusions were based on more convincing evidence than in the paper-based environment. They also noted that the binning capacity of TinkerPlots aided in the identification of outliers.

Watson and Donne (2009) studied learner understanding of beginning inference by comparing work in paper-based settings to work in TinkerPlots. Three different protocols were used to present the data to learners: two using paper-based techniques and a third using TinkerPlots to analyze the data. One of the paper-based techniques

presented learners with premade completed graphs that could not be manipulated, and the other presented learners with a set of cards with all the important aspects of each individual piece of data recorded on them. The students given the data cards were asked to create representations and form hypotheses. The representations they created included pictographs, pie graphs, bar graphs, line graphs, and scattergraphs. Different groups of participants, studied at different times, were used for the paper-based and TinkerPlots protocols. The two paper-based protocols used data from previous studies of student understanding that did not involve the use any software. The results of these studies are reported in Watson and Moritz (1999) and Chick and Watson (2001a, 2001b). The learners used for the TinkerPlots protocol had not been studied previously. The researchers found the major affordance of TinkerPlots as compared to paper-based tasks was the flexibility that it gave learners to create graphs based on their understanding of the data rather than having to conform to conventional graphs that were presented to them. The students working in TinkerPlots created a variety of different graphs that expressed their unique understanding of the data. Other affordances of TinkerPlots as compared to the paper-based protocols included speed of analysis. It was determined that the learners work in TinkerPlots was more efficient than the other two protocols.

Meletiou-Mavrotheris, Papparistodemou, and Stylianou (2009) conducted a study to explore how technology enhanced elementary teachers understanding of statistics in a professional development environment. Twelve elementary school teachers from Cyprus participated in the study. In the first stage teachers used only paper and pencil to compare the per capita income and mean life expectancies of people living in Europe to those in Cyprus. In the second stage teachers explored the

same data using TinkerPlots. The researchers found that in the paper-based environment teachers used numerical calculations and did not create any graphs using measures of central tendency to analyze the data. In the TinkerPlots environment, the teachers created graphs and used measures of variation along with measures of central tendency to explore relationships between the measures.

There are other studies using TinkerPlots that do not fall into the categories above. These studies have documented types of statistical analysis techniques that learners have used when working in TinkerPlots and also highlight the advantages of working in this environment. These studies illustrate how TinkerPlots can support statistical reasoning.

One study using TinkerPlots has documented the ease in which students were able to move flexibly between graphs and hypotheses. Fitzallen and Watson (2010) reported that this capability allowed 10 and 12 year olds using TinkerPlots to quickly test multiple hypotheses. Being able to click on a data point and view the values in the case card aided students in confirming their hypothesis.

Another study using TinkerPlots investigated the effect of students working on tasks set in meaningful contexts. Fitzallen and Watson (2011) reported that the understanding of context influenced 12 Year 5/6 participants' decisions and conclusions. Sometimes the students only made connections to the characteristics of graphs and sometimes their inferences were based on their understanding of context only. The researchers suggested that context be developed not only from interpreting contextual information from within the graph itself, but also developed from student's real-world life experiences.

Jointly, these studies support the argument that computer tools can enhance statistical reasoning. Ben-Zvi (2000) believes that statistics education should include technological tools because they provide opportunities for students to develop metacognitive capabilities. These tools do this because they permit active knowledge construction and they allow learners to reflect on what they see while using them. Today, technology tools have become indispensable tools for thinking and problem solving in statistics education where mundane exercises have been converted to higher order learning experiences.

## CHAPTER III

### RESEARCH DESIGN AND METHODOLOGY

Issues related to the development of students' thinking and understandings of statistical concepts were outlined in the review of literature. This study built upon and extended previously conducted research by adding the element of a computer-based software tool in order to determine the differences in student thinking in technological and nontechnological environments. In this section, I describe the participants, instrumentation, and procedures used in the study; provide a comprehensive description of the data that were collected during the course of the study; and describe how the data were analyzed.

#### **Research Design**

This study employed a qualitative research design, utilizing multiple case studies. According to Merriam (1998) a qualitative study “seeks to discover and understand phenomenon, a process, or the perspectives and world views of the people involved” (p. 11). The phenomenon under investigation here was students' reasoning about data in two distinct environments. Data were collected from several cases to broaden the scope of the investigation. These multiple points of view strengthened the findings of this study (Merriam). When more cases are included in a study, the range in similarities and contrasting cases will include more variation, thus, I was better able to understand the phenomenon under investigation. Also, interpretations of these data are likely to be more



compelling due to the variations present in multiple cases.

### **Participants**

The population of this study consisted of students in grades 6 and 7 in a university laboratory school. Students from various cultural and socioeconomic backgrounds comprised the demographics of the school population. In order to provide a wider range of thought processes, children from two different grade levels formed the population of this study. This scenario contributed to a broader, more balanced characterization of students' statistical thinking. Twelve students, six from each grade, were selected for case-study analysis.

A list of potential candidates was created from teacher recommendations based on judgment of students' levels of performance in mathematics and identified as those students most willing to verbalize their thoughts. Volunteers from students judged to have a medium or high level of mathematical performance were sought because these students were more likely to exhibit a broader range of thought processes from which more information about students' statistical thinking was observed. Students were assigned a partner to work with throughout the study. Research documents that when students use computer technology, small-group learning is more effective than individual learning in promoting greater individual academic achievement, group task performance, and positive attitudes toward learning (Lou, Abrami, & d'Apollonia, 2001). Pairs were comprised of students of the same grade level and selected based on teachers' recommendations of students they believed would work well together.

### **Pilot Study**

In November 2002 a pilot study was conducted for several reasons: (a) to

familiarize myself with the TinkerPlots software, (b) to determine how long it would take for students to come up to speed with using the software, (c) to develop the tasks to be used in the study, and (d) to observe student interaction with the software and their peers. The pilot study was carried out on three consecutive Fridays at Thomas Metcalf Elementary School on the campus of Illinois State University in Normal, Illinois. Two classes of fifth-grade students, 24 students in each class, participated in the study during their computer lab period. Each class period lasted 45 minutes. Students worked in pairs on the computer. The computer actions and movements on 3 computers in each class were recorded. Student generated work was collected at the end of each of these instructional sessions.

The student work and the computer recordings were analyzed to determine whether the tasks used were appropriate for the research study. This also proved beneficial in that it helped me modify and refine the tasks. Analysis of students' interaction with the computer and with their peers helped me further understand the benefits and limitations of using TinkerPlots to analyze students' statistical thinking.

The software was functional and worked well, without any major glitches. It challenged the students' thinking and created an environment where students were free to explore. The students were able to navigate in the software, and they enjoyed working with TinkerPlots. Students were successful in exploring and solving the tasks, demonstrating that these tasks would be appropriate for the study.

## **Instrumentation**

### **Tasks**

Students' thinking was examined as they engaged in two tasks. The tasks, based

on previous work on the statistical thinking of middle school students (Johnson & Hofbauer, 2002; Mooney, Langrall, Hofbauer, & Johnson, 2001) and the results of the pilot study, were used to assess the 12 middle school students' statistical thinking. One task was worked using the software TinkerPlots and the other task was worked using paper and pencil.

Both tasks required students to read, organize, represent, and analyze and interpret data. The primary focus of the tasks was to assess how students interpret and analyze data. The tasks were designed to elicit critical, creative, and statistical thinking. Tasks engaged students in comparing distributions and areas pertinent to comparing distributions including measures of central tendency, measures of variation, and the creation of representations.

In the task, Classical Music, the participants were asked to determine whether students studying for a math test while listening to classical music performed better than those who did not. In the task, data is presented to the students in an organized way but may need to be reorganized to complete the task (Figure 2 and Appendix B). Participants were required to work this statistical task using the software TinkerPlots.

#### Statistical Task: Study Habits

Mrs. Jones was talking to the students in one of her mathematics classes about an article she read. The article described a group of children who performed better on tests because they had listened to classical music while studying. After hearing this, some of her students planned to listen to classical music while they studied for their next math exam.

To see how well students performed on this test, Mrs. Jones had her students mark an 'X' next to their name on the test if they listened to classical music while studying. The results of the 80-point test are listed below.

Analyze the data to see if students who listened to classical music while studying performed better on the math test than students who did not listen to classical music

while studying.

1. How did the test scores for students who listened to classical music compare to the test scores of students who did not listen to classical music while studying?
2. What conclusion can you make about the effect of listening to classical music while studying?

Jason	63
Anna x	61
Lisa	48
Robert	57
Steven	70
Carla	72
Aaron x	68
Juan	73
Li	60
Terri	61
Casey	64
Shawn x	73
Kayla x	67
Portia	51
Derek	66
Barb	66
Marie	68
Phil x	78
Ian	56
Michael x	60
Linda	46
Cheryl	45
Stephanie x	65
Paul	56

Mary	54
Dana	55
Laura	80
Mia x	77
Nina	60
David	63
Sean x	40
Alan	68
Cindy x	66
Dan	71
Melissa	59
Diane	52
Edward	45
Jamal x	71
Kathy	60
Lee	74
Frances x	72
Latoya x	60
Gina	65
Ricky	52
Kaye	58
Ben	56
Chris	71
Toni	73

*Figure 2. Classical Music Task*

An isomorphic task was developed to assess students' statistical thinking while working in a paper and pencil format. In this task, participants are asked to determine whether students taking a Drivers' Education course performed better on their drivers' exam than those who did not. Data were presented to the students in tabular form and as a set of cards (Figure 3 and Appendix C).

### Drivers' Education

Mr. Williams, a high school Drivers' Education instructor at Jefferson High School in Missouri, requested some information from the Missouri Department of Motor Vehicles. This information showed the scores of students from Jefferson High School on the 80-point written portion of the test who took the drivers exam last year. Some of these students took the Drivers' Education course and some did not because Missouri does not require all students to take Drivers Education courses.

Mr. Williams would like for you to analyze the data to see if students who took Drivers' Education scored higher on the drivers' examination than those who did not take Drivers' Education. The students who did take Drivers' Education have an "x" next to their name on the list below.

Analyze the data to see if students who took Driver's Education performed better on the test than students who did not take Driver's Education.

1. How did the test scores for students who took Driver's Education compare to the test scores of students who did not take Driver's Education?
2. What conclusion can you make about the effect of taking Driver's Education?

Jacob	53	Megan	44
Amanda x	51	Deborah	45
Lori	38	Lindsey	70
Richard	47	Mya x	67
Scott	60	Nancy	50
Carrie	62	Dale	53
Adam x	58	Seth x	30
Julio	63	Alex	58
Ling	50	Carol x	56
Tammi	51	Dennis	61
Caleb	54	Monica	49
Shane x	63	Dawn	42
Katie x	57	Eric	35
Paige	41	Jovan x	61
Don	56	Kelly	50
Becky	56	Luke	64
Marcie	58	Frank x	62
Peter x	68	Latisha x	50
Ira	46	Gail	55
Matthew x	50	Ryan	42
Lynn	36	Kristin	48
Connie	35	Brett	46
Sarah x	55	Chad	61
Patrick	46	Tori	63

*Figure 3. Drivers' Education Task*

## **Research Procedure and Data Sources**

The implementation of the study was comprised of three major parts: one paper and pencil problem-solving session, two software training sessions, and one computer problem-solving session. This study utilized multiple sources of data: (a) video and audio tapes of students' problem-solving sessions, (b) recordings of computer actions and movements, (c) student artifacts, and (d) researcher observations and field notes. The wide range of data sources facilitated the triangulation of data and generated different perspectives that provided a comprehensive view of students' statistical thinking.

### **Software Training Sessions**

All participants were brought together for the software training sessions to learn to navigate in the software program TinkerPlots. Students worked in their assigned pairs throughout the two sessions in order to facilitate thought processes and generate interaction and communication between the students. The two software training sessions engaged students in activities that familiarized them with how to use TinkerPlots. On the first day, a modified version of the activity *Cats*, from the Investigations in Number, Data, and Space curriculum (Rubin, Mokros, & Friel, 1996) was introduced. In this activity students were asked to organize data, by creating graphical representations, and then use these representations to answer questions that compare characteristics of the 24 cats in the dataset.

In the second software training session, students continued to develop their skills with TinkerPlots while working with the *Cats* activity. Students were asked to present their solutions to the rest of the students in the training session, which was followed by a group discussion. I addressed issues concerning the software that were not covered in the

first training session, answered questions, and mediated the whole-class discussion. By the end of the two training sessions students were familiar enough with the TinkerPlots software to be able to effectively navigate in its environment.

### **Problem-solving Sessions**

I conducted the problem-solving sessions individually, with each pair of students. Each session – one for paper and pencil and one for TinkerPlots – was about 45 minutes to 1 hour in length. The session began with me reading the problem to the students aloud. The students were then asked to interpret and analyze the data. Once the pair completed the task, they were asked to share their work and discuss the meaning of their representations, analysis, and interpretations of the questions. I then asked questions that were designed to encourage students to elaborate on their thoughts and to reflect on the strategies of their partner. For example, I asked questions like the following:

1. Why did you draw it that way?
2. I'm not clear on how you got that. Can you say more?
3. How did you figure that out? Can you show me?
4. Do the graphs represent the same data? How can you tell?
5. What are the similarities and differences between the two graphs?
6. What do you think your partner drew the graph this way?

All sessions were audiotaped. The audio allowed me to critically review each session and to observe participants' interactions within the environment in which they were working. All discussion from the audio tapes was transcribed for analysis.

I systematically observed each problem-solving session and took handwritten notes of what was occurring. These data sources provided additional information

regarding the participants' statistical thinking and interaction with the environment within which they are working. In addition, I maintained a reflective journal in which thoughts on each session were kept.

**Paper and pencil task.** For the paper and pencil task the students were supplied with both the data set on paper and on separate cards that contained each respective piece of data. The paper and pencil sessions were audio and videotaped using a video camera. The video recordings preserved nonverbal communication as well as the sequences in which the students performed the tasks. All student work generated in the paper and pencil session was collected. The primary source of these artifacts was in the form of paper and pencil drawings generated while solving the statistical tasks.

**TinkerPlots task.** In the TinkerPlots problem-solving session, the task was isomorphic to the task the participants solved using paper and pencil. Students were supplied with both the data programmed into the computer software program TinkerPlots and also on paper. The data in TinkerPlots were presented to the participants in a random arrangement of 48 circles in differing shades of blue. Each circle represented a student and his or her score on the mathematics test. Each circle was filled with a shade of blue that reflected the student's test score. Darker shades of blue represented higher scores whereas lighter shades of blue represented lower scores. Refer to Appendix A for an image of the way the data were presented to the students.

Both audio and computer movements while working in TinkerPlots were recorded using a screen capture software program Camtasia Studio (version 1.0.1) by TechSmith. Camtasia stored this information on computer disk as .AVI files. More information regarding Camtasia Studio can be obtained at the website <http://www.techsmith.com>. The



screen captures preserved nonverbal communication as well as the sequences in which the students performed the tasks.

All student work generated in the TinkerPlots sessions were collected. The primary source of these artifacts was a data sheet on which the problem had been presented to the students. Not much had been recorded on the data sheets.

## **Data Analysis**

### **Data Analysis Overview**

I used qualitative methods based on Grounded Theory (Glaser & Strauss, 1967; Strauss & Corbin, 1990, 1998) to analyze the data collected. According to Charmez (2006), “grounded theory methods consist of systematic, yet flexible guidelines for collecting and analyzing qualitative data to construct theories ‘grounded’ in the data themselves”(p. 2). Grounded theory methods use inductive analysis to explore the data for themes and patterns. Both categories identified from *a priori* ideas based on existing research and categories that emerge from the data can be used to guide the analysis.

Grounded theory contributes insights “into existing or emerging concepts that may help to explain human behavior” (Yin, 2011, p. 8), which makes it suitable for the characterization of statistical thinking in technological and nontechnological environments. Birks and Mills (2011, p. 16) described grounded theory as appropriate when (a) little is known about the area of study, (b) generation of theory with explanatory power is a desired outcome and (c) an inherent process is imbedded in the research situation that is likely to be explicated by grounded theory methods. This study meets all of these criteria.

I used multiple techniques associated with grounded theory throughout the analysis of the data. Analysis of the data relied on a data reduction approach to determine patterns in the thinking of students'. Data reduction is a process that involves selection, simplification, abstraction, and transformation of raw data (Miles & Huberman, 1994). I sought to identify distinctive characteristics that appeared in the data. These characteristics included patterns and trends, as well as significant differences or commonalities of how children thought in a technological versus a paper and pencil environment. I used data reduction throughout the study which aided in analytic choices including the coding of data, summarization of patterns, and organizing the data into final conclusions.

To ensure consistency, I used the constant comparative method every time a passage was coded (Taylor & Gibbs, 2010). The passage that was coded was compared to ones that had already been coded allowing me to determine whether the passage or the previous ones had been appropriately coded. When new codes were identified, I checked all data that had been coded to see if there was any of the data that should be coded with the newly created codes.

To support the analysis of data, I used the software program Transana (version 2.52), developed by the Wisconsin Center for Education Research, University of Wisconsin-Madison. This software supports both text and digital video file formats. The handling of video is almost identical to working with text and allows for a direct comparison between the screen capture recordings and digital video files. All the audio from the problem-solving sessions were transcribed within Transana. Using Transana enabled me immediate access to all parts of the project, instant access to transcripts, the

ability to search for words or phrases, to create and retrieve coded sections of data, manage project and organize the data, and search for relationships between codes.

### **Data Analysis Phase I**

Data Analysis Phase I consisted of initial coding of the data pertaining to research question one as it related to critical and creative thinking and to research question two as it related to statistical thinking. During this phase of analysis the coding scheme was refined by adding, collapsing, expanding, and revising the thinking elements. I developed a coding framework that adequately characterized the data.

The process for coding the data and refining the codes during each of the cycles of coding was essentially the same. During each cycle, I read through the transcripts and marked (by circling and highlighting) everything that matched the definitions for the thinking process. I systematically scanned all transcripts for occurrences of each code. Words, phrases, and grouping of sentences were assigned codes. When new codes were identified, I checked all data that had been coded to see if there was any of the data that should be coded with the newly created codes. While I was coding I kept a master list of all the codes and how they were used. This helped me apply the codes to new segments of data.

I kept written notes during this coding process. In those memos I reflected on what I had learned about the data including the coding process, code choices, and relationships between codes. In addition, I recorded ideas that emerged, connections that were observed, and questions that I had.

**Critical and creative thinking.** For research question one, I modified Jonassen's Critical Thinking Model (2000) to analyze and evaluate the data. Jonassen's framework

provides a structure to evaluate and interpret the critical thinking of students in a technological environment that is consistent with constructivism. This model defines critical thinking skills as more than just a collection of separate skills but rather, an interactive system (Appendix D). This model describes a variety of thinking processes and how they are related to one another. Contained in this model are three basic components of critical thinking: content/basic thinking, critical thinking, and creative thinking. These three components are combined into complex thinking, which is the synthesis of content/basic, critical, and creative thinking. Complex thinking includes a combination of content/basic, complex, and critical thinking processes into multi-step larger action oriented processes. Jonassen used three components to compare and contrast the effects of different Mindtools – critical, creative and complex.

For my study, I used only the creative and critical components to analyze and interpret the data. I eliminated the complex component due to it being a synthesis of both the critical and complex thinking processes. Critical and creative thinking were made up of single thinking processes, whereas complex thinking involved many elements from both critical and creative that were formed into larger multi-step action oriented processes. I decided to focus on single thinking processes in this study. Since the data had not yet been analyzed for critical and creative thinking, analysis of complex thinking was not deemed feasible.

***Data coding cycle 1: critical and creative thinking.*** For the first cycle of coding I created tables for critical and creative thinking that listed the thinking processes of both components (Figure 4, Cycle 1). These tables were similar to Jonassen's (2000) tables to evaluate databases as Mindtools. Critical thinking involves the dynamic reorganization of

knowledge in meaningful and usable ways (Jonassen). The thinking processes that comprise critical thinking were broken up into three subheadings: evaluation, analyzing, and connecting. There are fifteen critical thinking processes under these three headings. Creative thinking requires going beyond accepted knowledge to generate new knowledge. The thinking processes that comprise critical thinking were broken up into three subheadings: synthesizing, imagining, and elaborating. There are fourteen thinking processes under these three headings. Each thinking process in the tables represented an individual code.

After the first cycle of coding, I found many similarities existed between critical and creative thinking elements. Jonassen (2000) noted that “many creative skills are closely tied to critical skills” (p. 28). I also found that that some of the thinking processes from both components, critical and creative, were related to one another. For example, hypothesizing and verifying were observed to occur in conjunction with one another. A pair would generate a hypothesis, then try to prove or show whether the hypothesis was correct. Therefore, I merged the elements of critical and creative thinking into one table (Figure 4, Cycle 2). Thinking processes were grouped under subheadings that provided a general description of the elements. Only the thinking processes were being used for coding, therefore the subheadings were eliminated.

In the new table (Figure 4, Cycle 2), I organized the thinking processes into groups that that were logically related due to similar themes. For example, concretizing, analogical thinking, and causal relationships were grouped together. Concretizing and analogical thinking had originally been components of creative thinking: concretizing under the subheading elaborating and analogical thinking under the subheading

synthesizing. Causal relationships had been a component of critical thinking under the subheading connecting. These thinking processes were grouped together because they each had to do with the use of context in the data.

Cycle 1		Cycle 2
Critical Thinking	Creative Thinking	Critical & Creative Thinking
<b>Evaluating</b>	<b>Elaborating</b>	
Assessing information	Expanding	Planning
Determining criteria	Modifying	Expanding
Prioritizing	Extending	Classifying
Recognizing fallacies	Shifting Categories	Finding sequences
Verifying	Concretizing	Recognizing patterns
		Comparing/contrasting
<b>Analyzing</b>	<b>Synthesizing</b>	
Recognizing patterns	Analogical thinking	Modifying
Classifying	Summarizing	Extending
Identifying assumptions	Hypothesizing	Shifting Categories
Identifying main ideas	Planning	
Finding sequences		Hypothesizing
		Verifying
<b>Connecting</b>	<b>Imagining</b>	Fluency
Comparing/contrasting	Fluency	
Logical thinking	Predicting	Concretizing
Inferring deductively	Speculating	Analogical thinking
Inferring inductively	Visualizing	Identifying causal relationships
Identifying causal relationships	Intuition	
		Identifying main ideas
		Logical thinking
		Inferring deductively
		Inferring inductively
		Summarizing
		Prioritizing
		Determining criteria
		Recognizing fallacies
		Identifying assumptions
		Assessing information
		Predicting
		Speculating
		Visualizing
		Intuition

Changes to Cycle 2  
 Combined Critical &  
 Creative Thinking  
 Eliminated subheading  
 Grouped by similar theme

Figure 4: Charts for Analysis Phases 1 through 5

Cycle 3	Cycle 4	Cycle 5
<b>Critical &amp; Creative Thinking</b>	<b>Critical &amp; Creative Thinking</b>	<b>Critical &amp; Creative Thinking</b>
Planning	Planning	Planning
Expanding	Expanding	Expanding
Classifying		
Finding sequences		
Recognizing patterns		
Comparing/contrasting	Comparing/contrasting	Comparing/contrasting
Modifying		
Hypothesizing	Hypothesizing	Hypothesizing
Verifying	Verifying	Analysis based observation
Fluency		Verifying
		Knowledge Building
Concretizing	Concretizing	Concretizing
Analogical thinking	Analogical thinking	Analogical thinking
Identifying causal relationships	Identifying causal relationships	Identifying causal relationships
Identifying main ideas	Identifying main ideas	Identifying main ideas
Inferring inductively	Inferring inductively	Inferring inductively
Summarizing	Summarizing	Summarizing

Changes to Cycle 3  
Deleted Thinking Elements

Changes to Cycle 4  
Deleted Thinking Elements

Changes to Cycle 5  
Added Two Thinking Elements

Figure 4: Charts for Analysis Phases 1 through 5



***Data coding cycle 2: critical and creative thinking.*** I coded the data for a second time using the table displayed in Cycle 2 of Figure 4. During this cycle of coding I did not find any evidence of the following elements of critical and creative thinking: identifying assumptions, inferring deductively, assessing information, determining criteria, prioritizing, recognizing fallacies, extending, and shifting categories. Different factors played a part in the absence of these elements, such as the question not being conducive to eliciting these types of thinking processes.

In addition, I found that the thinking element predicting was included within hypothesizing. Therefore, it was merged with hypothesizing and the thinking element predicting was eliminated from the table. Based on Jonassen's definitions of creative thinking, speculating and intuition were not observed. They were not recognizable in transcripts or apparent in the other data sources. Both speculating and intuition were eliminated from the table. However, later in the analysis process, speculating was determined to be the same as a thinking element that I discovered during the fourth data coding cycle. I also found that the thinking process logical reasoning was contained within the other critical, creative, and statistical thinking elements. It was a component within the other elements, not a separate element on its own.

Changes to the coding table were made based on what I had found in the second phase of coding. The thinking processes of identifying assumptions, inferring deductively, assessing information, determining criteria, prioritizing, recognizing fallacies, extending, shifting categories, predicting, speculating, intuition, and logical thinking were eliminated from the critical and creative coding table. The modified table is represented in Cycle 3 of Figure 4.

***Data coding cycle 3: critical and creative thinking.*** I coded the data for a third time using the table displayed in Cycle 3 of Figure 4. During this cycle of coding I modified several thinking processes. Classifying, sequencing, recognizing patterns, modifying, and fluency were found to be included as thinking elements of both critical and creative thinking, and statistical thinking. In critical and creative thinking these elements did not share any relationship with other thinking processes. However, these elements contributed to the process of creating a representation which played a major role in statistical thinking. Therefore, I made the decision to eliminate them as thinking processes in critical and creative thinking, and place them in statistical thinking.

Changes were made to critical and creative coding table based on these modifications. The thinking processes of classifying, sequencing, recognizing patterns, modifying, and fluency were moved to the statistical thinking coding table. The modified table is represented in Cycle 4 of Figure 4.

***Data coding cycle 4: critical and creative thinking.*** I coded the data for a fourth time using the table displayed in Cycle 4 of Figure 4. Two new thinking processes were identified. The first of these, an “analysis-based observation” was a variation of the thinking process, hypothesizing. These were statements that were valuable to the understanding and analysis of the data; however, did not fully meet the criteria of a hypothesis. While not direct answers to the question, these statements were analytical in nature. Later in the analysis process I determined that the thinking element speculating was the same as an analysis-based observation. Speculating had been deleted from the table in data coding cycle 2. Since the coding had been completed using the term analysis-based observation, I decided to leave it in the table rather than changing it back

to speculating.

The other new thinking process that was identified, “knowledge building,” was a variation of verifying. It was observed that sometimes after a pair of students generated a hypothesis, they did not attempt to verify it; rather, they just continued to analyze the data. Their goal was to learn more about the data, while keeping an open mind about the outcome. When the pair used knowledge building, in lieu of verifying, they continued analysis, without trying to prove if an answer to the question that they had generated was true. Instead, they just continued to analyze the data, or “build knowledge about it.”

The critical and creative coding was modified to reflect the addition of these newly identified thinking elements. Analysis-based observation was placed as a subcategory under hypothesizing and knowledge building was placed as a subcategory under verifying. The modified table is represented in Cycle 5 of Figure 4.

*Data coding cycle 5: critical and creative thinking.* I coded the data for a fifth time using the table displayed in Cycle 5 of Figure 4. I went back and applied the new codes found in Data Coding Phase 4, analysis-based observation and knowledge building, to the whole data set. I checked all data coded previous to the identification of these new codes to see if there was any of the data that should be coded with these newly created codes.

### **Statistical thinking.**

*Data coding cycle 1: statistical thinking.* In analyzing the data for research question 2, the codes were not predetermined. For the initial coding I read the transcripts line-by-line and looked for themes, topics, ideas, and concepts. I identified important words, phrases, sentences, and passages of text. This type of coding where themes and

categories are allowed to emerge from the data without any preconceived notions from the researcher is referred to as open coding (Birks & Mills, 2011; Lewins & Silver, 2007; Strauss & Corbin, 1990). These “grounded” codes emerge from the data as a researcher’s prejudices are set aside.

Several iterations of coding were performed during this phase of analysis to identify and isolate general categories that highlighted statistical thinking elements. During this process I summarized and consolidated topics that I had highlighted and circled and began to label them according to common characteristics. The themes that emerged from this first phase of analysis were representation, means, extreme values and outliers, proportional reasoning, and different group size.

***Data coding cycle 2: statistical thinking.*** Cycle two required a back-and-forth analysis with the critical and creative thinking elements as codes were added, deleted or combined. Several thinking processes were moved from Jonassen’s Critical Thinking Model (2000) and added to the statistical thinking elements. Classifying, sequencing, recognizing patterns, modifying, and fluency were included as both elements of critical and creative thinking, and statistical thinking. Creating a representation of a data set plays a significant role in statistical thinking. These elements contribute to this process. In creative and critical thinking these elements did not share any relationship with other thinking processes. Therefore, I made the decision to eliminate them from critical and creative thinking, and place them in statistical thinking as a subcategory of representation.

To complement the themes that I identified during the first cycle of coding, the work of Makar and Confrey’s (2004b) research “Variation-Talk”: Articulating Meaning

in Statistics” helped me identify and frame the statistical thinking element variation. Makar and Confrey’s study was designed to gain insight into the ways in which prospective teachers expressed notions of variation in comparing data distributions in a relevant context. They identified standard and nonstandard ways participants expressed variation within the data (Appendix E). The task given to the teachers asked them to compare the improvement of test scores between two groups of students. Makar and Confrey’s observations and findings were a good match with my study. Participants in both studies were asked to compare two distributions of data.

Makar and Confrey (2004b) described conventional standard statistical language as proportion or number improved, means, outlier and extreme values, shape, standard deviation, and range. Nonstandard statistical talk was called “variation-talk” and classified into four categories: spread, low-middle-high, modal clump and distribution chunk. I noticed that much of the discussion of the participants in my study was nonstandard and contained nuances that were not captured by the term variation only. There was overlap in the themes that I had identified during my first phase of coding, and those identified by Maker and Confrey.

The themes that I had identified in the first cycle of coding were merged with the themes relating to variation that and Confrey (2004b) had identified. Three of the original themes that I had identified in the first cycle of coding, representation, proportional reasoning, and different group size, that were not included in the Makar and Confrey study, were included in my final framework. Two other themes, means and outliers were included in both my first cycle of coding and Makar and Confrey’s study were included in the final framework. In addition, several thinking elements identified

from the Makar and Confrey study, shape, range, spread, modal clump and distribution chunk, were included in the final framework.

Several elements identified by Makar and Confrey (2004b) were not included in the final framework. Standard deviation was not included because it was beyond the scope of knowledge of middle school students. During coding, low-middle-high was not observed and therefore not included in the final framework as well. Finally, proportion or number improved was not included in my coding framework. It was part of answering the questions being asked and was demonstrated by every pair.

***Data coding cycle 3: statistical thinking.*** Based on the first two phases of analysis I created a table of statistical thinking elements (Figure 5). Each element in the table represented an individual code. I applied all of the codes the whole data set. The passage that was coded was compared to ones that had already been coded allowing me to determine whether the passage or the previous ones had been appropriately coded. When new codes were identified, I checked all data that had been coded to see if there was any of the data that should be coded with the newly created codes. Revisions to the coding of the data set were made as necessary.

Statistical Thinking
Representation
Fluency
Modification
Sequencing
Classifying
Proportional Reasoning
Means
Outliers and Extreme Values
Shape
Range
Spread (cluster, clumped, trend)
Modal Clump
Distribution Chunk
Different in Group Size

*Figure 5: Elements of Statistical Thinking*

## **Data Analysis Phase 2**

After initial coding was completed in Data Analysis Phase I, tables were created that chronologically listed how each session was coded for critical, creative, and statistical thinking. A brief description of how the pair used each code was included in the table.

Based on Data Analysis Phase I, I wrote characterizations for each session. These characterizations interpreted each pair's work in a way that characterized their actions and understanding of critical, creative, and statistical thinking. My guiding question during this phase of analysis was "What did the analysis indicate about the students' critical and creative thinking, and statistical reasoning?" I interpreted what I had found in Data Analysis Phase 1 and explained what defined the students' thinking and actions on the tasks. I focused on explaining and interpreting the elements of critical, creative, and statistical reasoning that each pair exhibited.

After the characterizations for each session were completed, I interpreted differences in thinking between the two environments for each pair. My guiding question during this phase of analysis was “What did the analysis indicate about differences in the pairs’ thinking that were evident in the two environments?” The results of the initial coding from Data Analysis Phase 1, the chronological lists of codes, and the characterizations for paper and pencil and TinkerPlots were compared for each pair. These sources helped me describe patterns and identify themes and trends across both environments. Differences in a pair’s thinking emerged across both environments.

Based on this analysis, a summary called “Differences in Thinking When Working in the Two Mediums” was created for each pair. These summaries were my interpretation and description of how the pairs’ thinking differed in the two environments.

### **Data Analysis Phase 3**

For the third phase of analysis, I developed cross-case analysis summaries for each environment. During this phase of analysis, the characterizations and chronological lists of codes for each environment were examined. These sources helped me look for patterns, themes, and trends within each environment. I applied intermediate coding to the characterizations for each environment using the coding frameworks for critical, creative and statistical thinking developed in Data Analysis Phase I. During this process I became aware of broader patterns in the data. Relationships between the codes were found and categories emerged. From this coding commonalities and differences in the way pairs approached the tasks and their understanding of critical, creative, and statistical thinking were identified.



## Summary

The purpose of this study was to gain a better understanding of how students think in a technological environment. It employed a qualitative research design, utilizing multiple case studies. The population consisted of students in grades 6 and 7 in a university laboratory school.

Students' thinking was examined as they engaged in two tasks. One task was worked using the TinkerPlots software and the other task was worked using paper and pencil. The tasks were designed to elicit critical, creative, and statistical thinking. These tasks engaged students in comparing distributions and areas pertinent to comparing distributions including measures of central tendency, measures of variation, and the creation of representations.

Three phases of analysis were conducted. During the first phase, I developed a coding framework to adequately characterize the data. In the second phase of analysis, I developed characterizations for each session and compiled differences in thinking for each pair across both mediums. For the third phase of analysis, I developed cross-case analysis summaries for each environment.

## CHAPTER IV

### FINDINGS

This chapter presents an analysis of the data that was collected. The following questions were addressed:

1. How does the critical and creative thinking of middle school students using the software program TinkerPlots differ from their thinking while using a traditional paper and pencil format?
2. How does the statistical thinking of middle school students who use the software program TinkerPlots differ from their statistical thinking while using a traditional paper and pencil format?

This chapter is divided into two major sections: findings by case and findings of cross-case analysis. In the findings by case section, analysis of the six pairs of students is presented. The first section reports aspects of critical and creative thinking that are common across all pairs. The next six sections, one for each pair, contain (a) a summary of the paper and pencil session, (b) characterization of the paper and pencil session, (c) a summary of the TinkerPlots session, (d) a characterization of the TinkerPlots session, and (e) differences in thinking when working in the two mediums. The findings of cross-case analysis section is subdivided into (a) the cross-case analysis of the paper and pencil task, (b) the cross-case analysis of the TinkerPlots task, and (c) the summary of findings.

## **Findings by Case**

### **Aspects of Critical and Creative Thinking Common Across Pairs**

All pairs demonstrated the following elements of critical and creative thinking in both environments: planning, expanding, comparing/contrasting, and summarizing. Two other elements—classifying and identifying main ideas—were demonstrated by all but one pair, 6C. Use of these elements by each pair in both environments was essentially the same. They were interspersed throughout the entire session and were necessary in order to answer the question being asked. These elements were too pervasive and embedded in what the students were doing to be able to code them in a meaningful way. Therefore, they will be summarized here across all pairs in both environments.

None of the student pairs devised any formal plans or schemes for analyzing the data in either the paper and pencil or TinkerPlots environment. They did not establish a methodology or framework, or plan of action up-front. Rather, they planned as they went, and their desired goal was formed in the process. This allowed them the flexibility to base their next step on what they had learned from the previous one. Sometimes a change was made deliberately, through discussion of what to do next. At other times, planning of the next step was made “on the fly.”

Every pair in both environments “expanded” on the data. By working through the task the pairs increased the scope of the body of information they had been given. Expanding included all aspects of analyzing the data, such as adding details, creating representations, and hypothesizing about the conclusion. Expanding was a necessity in order to answer the question being asked.

The compare and contrast elements of critical and creative thinking were inherent

in the tasks used in this study. In each task, students were presented two data sets and were asked to compare them. It was essentially impossible to engage in these tasks and not compare and contrast the data in some way. Thus, it is not surprising that all student pairs exhibited this aspect of critical and creative thinking.

In both environments, every pair inferred inductively. After analyzing the data, they developed explanations, predictions, and hypotheses about the questions that were asked. Every pair developed theories about the data and drew a conclusion based on them.

Summarizing was demonstrated by every pair in both environments. Every pair was able to produce a succinct form of the material. They were able to identify the basic structure of the data and express it in a more consolidated form than the original. Summarizing was pervasive throughout the session and a necessity in order to answer the question being asked. Students summarized the data by representing it in various ways, such as in tables, graphs, and as averages. Also, I considered their hypotheses and conclusions as summary statements of their analyses.

Classifying was demonstrated by all pairs in both environments except pair 6C in paper and pencil. All except this one case, every pair classified the information they were given. In paper and pencil they sorted the data cards into two groups – students who took driver's education and students who did not take drivers education. In TinkerPlots they sorted the data into two groups – students who listened to classical music and students who did not listen to classical music.

Every pair identified the main ideas of the question in both environments, with the exception of pair 6C in paper and pencil. They recognized the two groups of students

were different sizes and that this would affect their analysis.

## **Pair 6A**

**Summary of paper and pencil session.** Pair 6A was comprised of two sixth-grade students, Amy and Heather. The task they worked on in paper and pencil was called Drivers' Education (see Appendix C). In this task, participants are asked to determine whether students taking a drivers' education course performed better on their drivers' exam than those who did not. Data were presented to the students in tabular form and as a set of cards.

Amy began the session by sorting the cards into two piles: students that took Drivers' Education and students that did not take Drivers' Education. The pair then ordered the data cards in each pile from smallest to greatest, according to test score. Their initial discussion is presented in Episode 1.

### Episode 1

- 1 A: Most of them didn't take drivers ed. ... The numbers are not even.
- 2 H: So far the average is in the 50s for taking.
- 3 A: That's good.
- 4 H: There are some that are an exception. And there are some that are, well ... that obviously did better
- 5 A: One of them got perfect that didn't take the drivers' ed. test.
- 6 H: Seriously? Lindsey?
- 7 A: ...would take if take out the outliers. I am going to put them in order.
- 8 H: Ok
- 9 A: Take out the outliers see what the range is around and then see which ones are closer and stuff like that.
- 10 H: Ok.
- 11 H: Only, the total is 13 people that took the driver's ed. course out of 48. So 13 people out of 48 people took the course.
- 12 A: 13 people out of 48 people
- 13 H: took the course

While working to organize and sort the data, Amy and Heather discussed observations they made. For example, they approximated an average for the students who took Drivers' Education. Heather stated that "So far the average is in the 50s for taking." They noticed exceptions to this average, including high and low scores. For example, one of the students who did not take the driver's education course received a perfect score on the test. They recognized that fewer students took Drivers' Education: 13 out of 48 students took the course whereas 35 out of 48 did not. And there was also a discussion about removing the outliers and looking at the range of the scores.

After this initial sorting, organization, and discussion of the data, Amy and Heather decided to represent the data by creating a back-to-back stemplot (Figure 6). They did not state why they chose to create this type of graph. On one side of the graph students who took Drivers' Education were represented while the other side showed students who did not take Drivers' Education. Heather drew the frame of the back-to-back stemplot and put the students who took Drivers' Education on the graph while Amy filled in the students who did not take Drivers' Education.

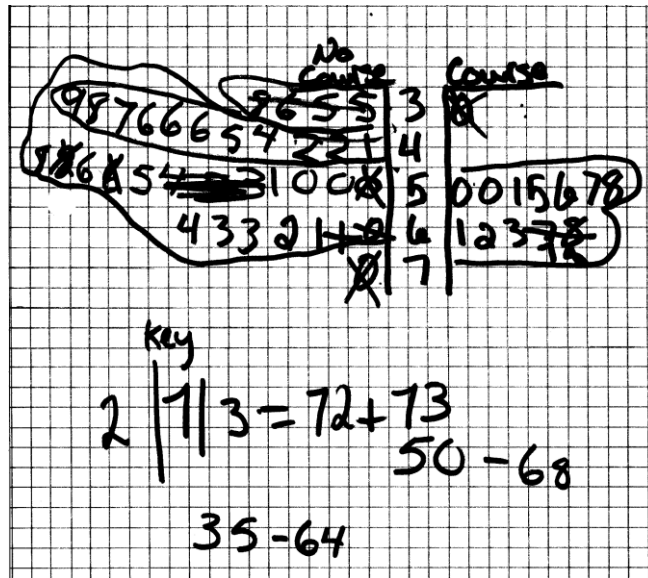


Figure 6. Back-to-Back Stemplot

Episode 2 illustrates how Amy and Heather analyzed the back-to-back stemplot they constructed.

Episode 2

- 36 A: There were a lot of people in the forties who didn't take the course. It looks right now like the people who took the course did better.
- 37 H: Scored higher.
- 38 A: But that's just because I haven't put on the fifties yet.
- 39 H: But how can you tell because there are so many more people that took, did not take the course. But the people that did not take the course ...
- 40 A: This is the average. (Amy circles fifties and sixties for the students who took the course.)
- 41 H: So should uh yea
- 42 A: ... for the people who um, this is an outlier. 50 – 68 (Amy crosses out the lowest score for students that took Drivers' Education, Seth, 30. Then writes 50 – 68 on paper, the range of scores for students that took Driver's Education, without the outlier of 30.)
- 43 H: And then really the only outlier is 70 for that. But... (Heather circles the 40s, 50s, and 60s for the students who did not take Drivers' Education. Amy crosses off the highest score for the students who did not take Drivers' Education, 70).
- 44 A: 35 to 64. Ok, that's, it, the people who took the course did better. (Amy writes 35 - 64 on paper, the range of scores for the students who did not take Drivers' Education.)
- 46 H: But it, people in the fifties, those people had ... (Heather points to students who took Drivers' Education.)

47 A: There are more fifties over here. (*Amy points to the students who did not take Drivers' Education.*)

48 H: True. What about sixties?

49 A: Had more people...

While creating the back-to-back stemplot, Amy hypothesized that students who took the driver's education course performed better on the test (line 36). This was the pair's first hypothesis and it was questioned by Heather (line 39), which prompted the pair to continue to analyze the graph and verify Amy's hypothesis. In the process of their analysis, Amy noted that there were more students in the 50s and 60s that took the course, calling this the "average" for that group (line 40). After deleting what they called the outliers for both groups, Amy compared the ranges for both groups of students and determined that the range for the students who took Drivers' Education was higher than the range for the students who did not take Drivers' Education (lines 42-44). The range for the students who took drivers education was 35 – 64 and the range for the students who did not take drivers education was 50 – 68. With the graph completed, Amy was able to see the entire picture the data painted and stated that "the people who took the course did better" (line 44). This was a restatement of the pair's original hypothesis. Heather was still perplexed with Amy's hypothesis that the students who took Drivers' Education did better on the test than the students who did not take Drivers' Education. She continued to question Amy's conclusion (lines 46-49). Amy recognized that Heather's frustration stemmed from the fact that there were different numbers of students in each group. Heather's questioning prompted Amy to begin modifying the stemplot.

Because the subsequent dialog between the students is difficult to follow, I have not included the actual transcript but describe what occurred. In an effort to help Heather



deal with the unequal group sizes, Amy began to delete scores from the larger group, the students who did not take drivers' education, in order to make the groups equivalent in size. The pair referred to this process of elimination as "skimming-off." To delete a test score, the pair would physically cross it off of the back-to-back stemplot with a pencil.

Initially, Amy and Heather crossed-off the test scores of students in the larger group, those who did not take Drivers' Education, but only if the other, smaller group of students, those who took Drivers' Education, had a matching score. For example, both groups of students had scores of 56 and 58 so these test scores were crossed-off from the larger group. Following this procedure for a while the pair crossed-off scores but then for no apparent reason they altered their strategy. What was once a well-reasoned and deliberate affair turned into a free-for-all, that is, test scores were crossed-off from both groups.

Amy deleted scores from the larger group if there was *no* matching score in the other, smaller group, which held with the original plan the pair had constructed. For example, the students in the larger group had test scores of 54, 53, 53, and 50 that the students in the smaller group did not have. Therefore, these scores were crossed-off for the larger group. What greatly confused the process was that Heather began to cross-off test scores in the smaller group if there was no matching score in the larger group. For example, there were test scores of 67 and 68 in the smaller group, those who took Drivers' Education, but these test scores could not be found in the other, larger group. Heather crossed-off these scores from the list of scores for the smaller group! Once the group sizes were approximately equivalent, Heather was comfortable enough with the data to make a comparison. Episode 3 illustrates how Amy and Heather concluded the

session by responding to the questions being asked.

### Episode 3

61 H: How did the test scores compare?

62 A: Yea. (*Amy begins writing response to question 1.*)

63 H: Well, um, um, the people that took the course they scored more in the fifties and sixties range and only had one thirty, one super low score, but, the people that did not take the course they had more in the thirty fives and forties which the people then... One person who did not take the course scored higher and one person who took the course scored lower.

64 A: The test scores compared to the people who took the course had around 50 to 64 about and the people who didn't take the course had more 30 through 40 so basically the people who took Drivers Ed. did better.

65 H: Taking Drivers Ed. would help

66 A: Yes

67 H: And the process of taking Drivers Ed.

68 A: The conclusion we can make about taking Drivers Ed. is that it would help your test score.

The pair was able to identify that the students who did not take the course had many students who scored in the 30s and 40s (lines 63 - 64). While there were not as many students it that took the course, most of them scored in the 50s and 60s. Without being prompted by the researcher the pair concluded the session by responding to the second question, "What conclusion can you make about the effect of taking Drivers' Education?" The pair both agreed that on the following response: "The conclusion we can make about taking Driver's Ed. is that it would help your test score" (lines 65 – 68).

**Characterization of Pair 6A paper and pencil session.** Table 1 is a chronological listing of how pair 6A's session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

6A Critical/Creative Thinking	6A Statistical Thinking	Description
	Classifying	Sorts data cards into two piles
	Different Group Size	"Numbers not even."
	Average	Uses term "average"
	Modal Clump	Refers to scores in 50s
	Extreme/Outlier	Refers to perfect score
	Sequencing	Orders data cards
	Extreme/Outlier	Uses word "outlier"
	Range	Uses word "range"
	Different Group Size	Recognizes fewer students took DE ?
	Sequencing	Continues to order data cards for no DE group
	Average	"I think 50 was the mean."
	Representation	Back-to-back stemplot
	Extreme/outlier	"Peter scored the highest."
Hypothesis		Looks like people who took the course did better.
Verification of Hypothesis	Different Group Size	So many more people did not take course.
	Average	Uses word "average."
	Modal Clump	Circles data points in 50s and 60s
	Extreme/Outlier	Uses word "outlier." Deletes low score from graph.
	Range	Refers to range as "50 to 68"
	Extreme/Outlier	Uses word "outlier." Deletes high score from graph.
	Range	Refers to range "35 to 64"
	Distribution Chunk	Circles data points in 40s, 50s, 60s
Restatement of Hypothesis		"The people who took the course did better."
	Distribution Chunk	"There are more fifties over here."
	Different Group Size	Says there are more people that took DE
Verification of Restatement of Hypothesis	Modifying Representation	Crosses scores off stemplot
	Distribution Chunk	"they got more in the 30s", "these people got a lot of 40s."
	Extreme/outlier	Uses word "outlier"
	Range	Refers to data in the range from 50 – 64
	Distribution chunk	"People who didn't take Driver's Ed. had more 30 through 40"
Conclusion		"It would help your test score."

*Table 1.* Chronological List of Codes for Pair 6A Paper and Pencil Session

Amy and Heather used statistical thinking elements to help them organize the data. They classified and sequenced the data, which helped them create their representation; a back-to-back stemplot. The statistical techniques the pair used in the analysis of the back-to-back stemplot were means, extreme/outliers, range, distribution

chunk, modal clump, and recognition of different group size. In addition, the critical and creative thinking elements used in the analysis were hypothesizing and verification.

The girls' work on this task can be characterized by an analysis cycle that included analyzing the graph they created, stating a hypothesis about which group of students did better on the Driver's Education test, and attempting to verify the hypothesis. This cycle was repeated twice during the session. It is noteworthy that in both cycles the verification was prompted by Heather's difficulty in dealing with the fact that the data sets did not contain the same number of data points.

The way the pair attempted to handle the different sized data sets prompted another characteristic aspect of their work – an invented “skimming-off” strategy for equalizing the data sets. The pair deleted outliers and extreme scores from the graph in a process they referred to as “skimming-off.” This technique was unconventional and not implemented systematically. The original back-to-back stemplot was an accurate representation of the data; the modified version was flawed due to the haphazard application of this technique. This is significant because while the pair did attend to the difference in group sizes, they ignored some features of the distributions when making comparisons. Outliers and extremes were no longer a part of the distributions and the size of the data sets was now equal. They did arrive at a valid conclusion; however, this was based on a graph that was only partially representative of the data.

Amy and Heather's work on this task is also characterized by their use of “approximate averages,” modal clumps, and distribution chunks. The back-to-back stemplot allowed Amy and Heather to visualize the distribution of each data set arranged from least to greatest. They were able to view the placement of both sets of data in

relation to each other. These visualizations aided the pair in developing what I called “approximate averages” for both sets of data. An approximate average was a visual observation as opposed to a mathematical calculation. The pair used the term “average,” but did not calculate the arithmetical mean. What they meant by “average” was that they could see a pattern in the data that the majority of the students in a group scored in a particular range. The “averages” identified were modal clumps, a range of scores at the heart of the distribution of students who took drivers education. This represents a flexible use of the range as an interval of values rather than a measure.

After the graph was modified and the group sizes were approximately equal, the modal clumps that the pair had identified earlier in the session shifted to distribution chunks. In contrast to modal clumps, distribution chunks are groupings of data not in the center of the distribution. Recognition of distribution chunks involves discussion about groupings of the data including phrases such as “the bulk of them, the main group, the big chunk, and the majority” (Makar & Confrey, 2005).

In summary, Amy and Heather’s work in the paper and pencil environment was characterized by repeating the analysis cycle twice during the session, attempting to equalize the data sets using an unconventional technique, and using the statistical thinking elements “approximate averages,” modal clumps, and distribution chunks in their analysis.

**Summary of Pair 6A TinkerPlots session.** The second task Amy and Heather completed required them to determine whether students who studied for a mathematics test while listening to classical music performed better than those who did not (see Appendix B). The data in TinkerPlots were presented to the participants in a random

arrangement of 48 circles in differing shades of blue. Each circle represented a student and his or her score on the mathematics test. Each circle was filled with a shade of blue that reflected the student's test score. Darker shades of blue represented higher scores whereas lighter shades of blue represented lower scores. Refer to Appendix A for an image of the way the data were presented to the students.

The session began with Heather controlling the mouse, while Amy gave her instructions on how to organize the data. Heather asked Amy if she thought the data should be sorted into columns. Amy said yes and Heather separated the data into two columns labeled "yes" and "no" based on whether the students had listened to classical music or not. They manipulated the data by clicking on the tool to create circle graphs, and next clicked on the tool to order the scores from least to greatest. This resulted in the construction of two side-by-side circle graphs (Figure 7). One circle graph consisted of students who had listened to classical music and the other of students who had not listened to classical music. Both graphs were ordered by test score. At this point, the pair had enough information to begin analyzing the data.

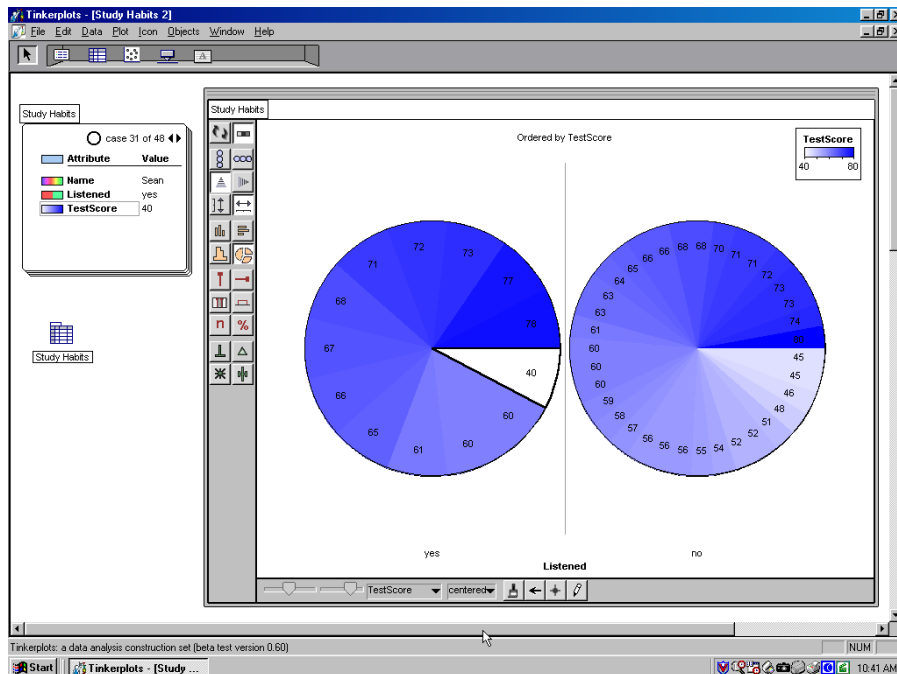


Figure 7. Side-by-Side Circle Graphs

Episode 1 illustrates how Amy and Heather initially analyzed the circle graphs they constructed.

### Episode 1

- 1 A: Is that a good grade or is that a bad grade? (*Referring to the student Sean who listened to classical music yet scored the lowest overall, 40 points out of 80. Sean's test score is represented by the lightest section in the circle graph of students who listened to classical music.*)
- 2 H: That is a bad grade that is 50. Um, he did listen...
- 3 A: Put up the key.
- 4 H: That, ok.
- 5 A: Let's see...
- 6 H: How do I...
- 7 A: There.
- 8 H: There we go. Ok. Sean did listen to music.
- 9 A: He's an outlier. He is 20 away from the other ones. He must have just been listening to classical music and not studying.
- 10 H: Probably.

Their discussion focused on Sean, a student in the group that listened to classical music, who received the lowest test score overall. Amy referred to Sean as an outlier because his

score was 20 points lower than the other students who listened to classical music. This was very apparent on the graph since the color of the section representing Sean's test score contrasted sharply against the other sections of the graph. In line 2 when Heather referred to Sean's "bad" grade as being "50," she had essentially interpreted the raw test score of 40 points as 50%. Since the pair was familiar with grades in their everyday school experience, this put the test scores into a context they could understand. In line 9, Amy offered an explanation for why Sean scored so low—that he was listening to the music but not studying.

The pair altered Figure 1 by splitting the two circle graphs into multiple smaller circle graphs in two columns (Figure 8). The vertical axis separated the circle graphs by whether classical music was listened to, or not. In each column the circle graphs were placed in bins of five points each according to test scores (Figure 8). Each circle graph within a bin contained scores within the five point range.



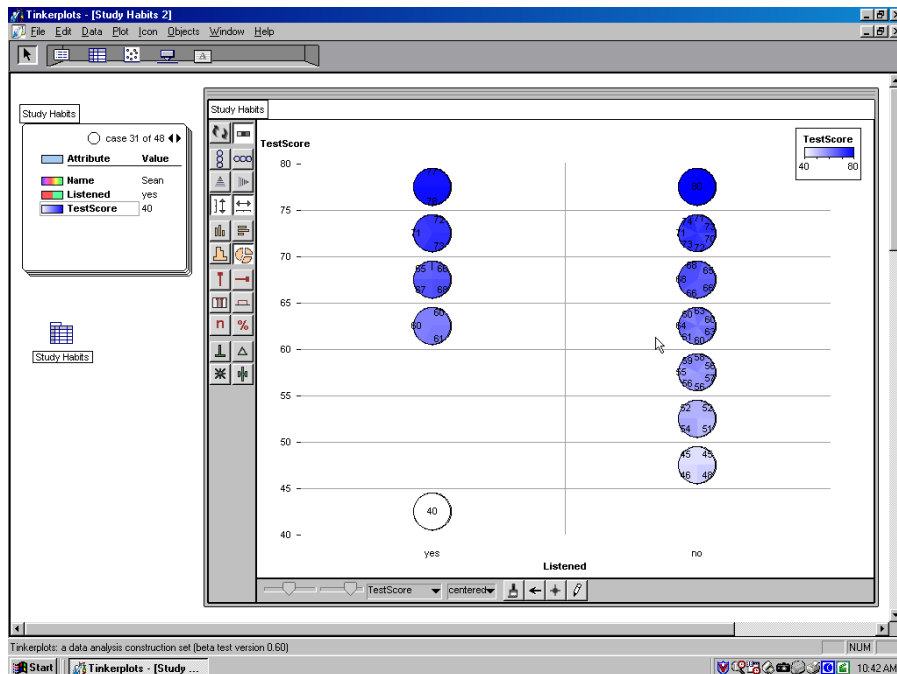


Figure 8. Multiple Circle Graph

The pair’s analysis of this graph was brief. Heather noticed that there were fewer students who listened to classical music. Episode 2 illustrates how Amy and Heather analyzed the multiple circle graph they created.

Episode 2

20 A: Ok, um, there’s more circles lower.

21 H: She’s but, yet not as many students.

22 A: I think it helps. We shouldn’t make any conclusion yet. Let’s put it into a bar graph.

23 H: Ok.

Heather noticed that there were fewer students who listened to classical music. Amy’s statement that “there’s more circles lower” referred to the three circles on the left-hand side of the graph that represented test scores of less than 60. Heather’s comment about “not as many students” was not acted upon and Amy made the pair’s first hypothesis that listening to classical music helped students score higher on the test. The pair exhibited caution and decided to analyze the data further before accepting Amy’s initial hypothesis

as their final conclusion. Everything the pair did during the remainder of the session was for the purpose of verifying their initial hypothesis.

Amy and Heather continued their analysis of the data by creating a different type of graph; a bar graph that was ordered by test score and colored to show whether or not the students listened to classical music (Figure 9). The students that listened to classical music were represented with the color green and the students that did not listen to classical music were represented with the color red.

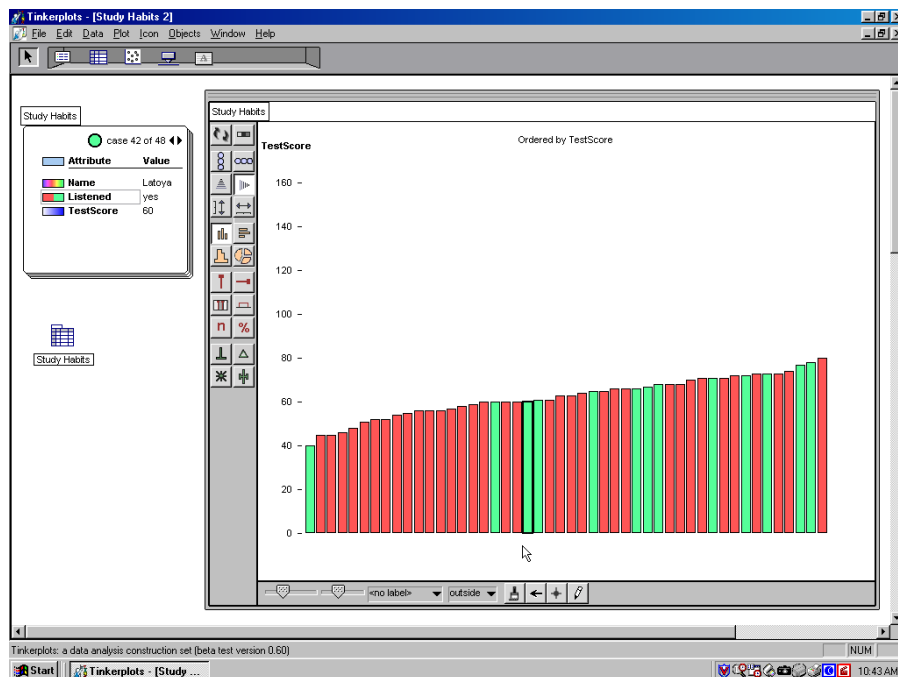


Figure 9. Bar Graph

In Episode 3, the girls interpret the bar graph to verify their hypothesis that listening to classical music helped students score better on the test.

### Episode 3

27 H: Like this? (*Heather asking Amy if she should leave the graph like Figure 9.*)

28 A: Yea. No, actually this is good. It shows it helps...

29 H: The test scores.

30 A: Yea.

31 H: Yea. Put the key up there.

- 32 A: I think the majority of students who didn't listen to classical music were down here. (*Toward the left side of the graph represented by lower scores.*)
- 33 H: Yea, one though did score higher. (*Referring to the tallest, rightmost bar.*)
- 34 A: Yea, but she's just an exceptional student.
- 35 H: Laura got ... um...
- 36 A: Um, how it compared? The people who did listen to classical music were up higher. They were in the sixties and seventies. While most of the people who didn't listen were down in forties and fifties.
- 37 H: Forties through sixties.
- 38 A: Yea. 'cos there is that big clump of people right there. (*Referring to the large group of consecutive red bars at the left side of the graph.*)
- 39 H: Um, and there is one outlier for the group that did listen to classical music who only scored forty on his test.
- 40 A: He scored fifty percent.
- 44 A: Ok. What conclusion can you make about the effect of..., um, would it help you score better test, um, it might help you score higher than you would normally score, but you shouldn't get so absorbed in the classical music that you don't study.
- 45 H: So you shouldn't get so absorbed in the music that you don't study.
- 46 A: Yeah, you should have it low like those people. (*Referring to the green bars toward the right side of Figure 9.*)
- 47 H: Like that person? (*Referring to the rightmost green bar on Figure 9.*)
- 48 A: Yeah, they got good test scores, they were better.

After the pair created the single bar graph (Figure 9), Amy immediately noted “Yeah, No, actually this is good. It shows it helps.” What she meant by this statement was that the graph showed that the students who listened to classical music scored higher on the test than those students who did not listen. This was a restatement of the hypothesis presented earlier in the session (episode 2, line 22).

The pair continued to analyze Figure 9. Heather and Amy focused on the colors and heights of the bars and identified the locations of “chunks” of red and green bars. They noticed that there was a large clump of consecutive red bars (i.e., students who did not listen to classical music) toward the left side of the graph, which represented lower test scores (lines 32, 38). Also, they noticed there was a higher concentration of green bars toward the right side of the graph; that is, students who had listened to classical music

had higher test scores. They noted that students who listened to the classical scored in the sixties and seventies, while most of the students who did not listen to classical music scored in the forties and fifties (lines 36-37). Heather pointed out that although the majority of students who did not listen to classical music had low scores, the student with the highest overall test score did not listen to classical music. Amy attempted to explain this aspect of the data by referring to the student as being “exceptional” (lines 33-35). She also mentioned that the lowest score was from a student who listened to classical music, despite the fact that the other students who listened to classical music had higher test scores (line 39). Again, she attempted to explain this situation by implying that the student had become so absorbed in the music that he failed to actually study.

After the pair’s conclusion (lines 44-48), they paused for a moment, and then without explanation, dragged the green bars in Figure 9 to the left side of the graph. This action split the data into two bar graphs (Figure 10). The bar graph on the left represented students that listed to classical music and the graph on the right represented students that did not listen to classical music.

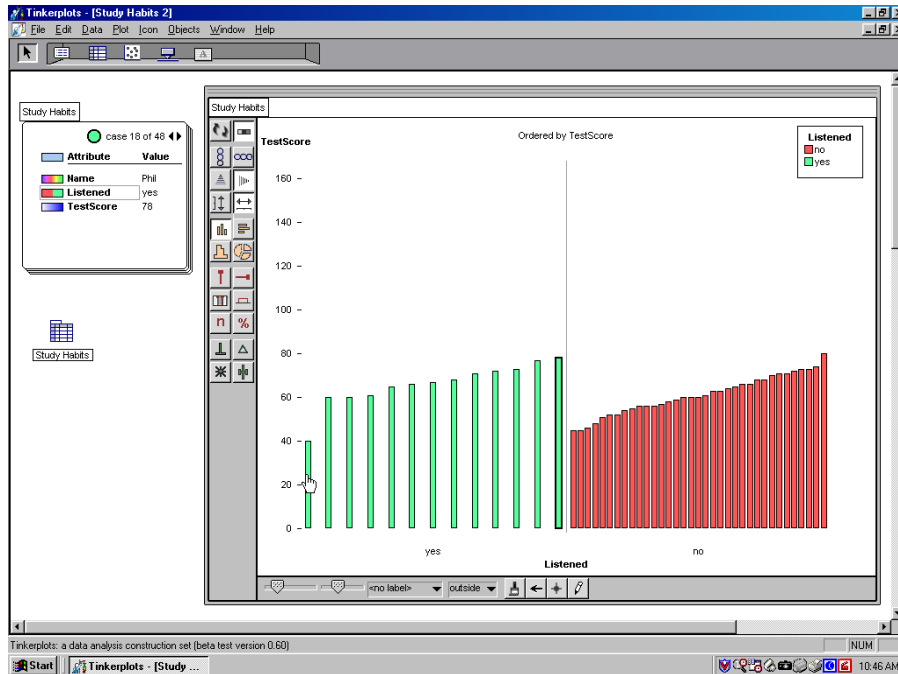


Figure 10. Side-by-Side Bar Graph

Episode 4 illustrates how Amy and Heather analyzed the side-by-side bar graph they constructed.

#### Episode 4

57 A: Yea these people did most in

58 H: The seventies, 72, seventy...

59 A: Yea most of them are in the seventies. Or in the high sixties.

60 H: Yea, from about sixty

61 A: 60 to 78

62 H: With the exception of the outlier who scored 40.

63 A: This is also a fact to prove our evidence. Oh, let's put the median up. Ok this, this is also a fact to prove our evidence. The...is this the median?

64 H: That is the median.

65 A: This is the median and it is 60 for the people who didn't listen to classical music while they studied, and it's 70 for the people who did listen to classical music. Another fact to prove our evidence. What's that?

66 H: The mode.

67 A: Ok, it's the mode.

68 H: Um, the mode is not that much different.

69 A: But it is still different.

70 H: Yea, like

71 A: The mode in this class is 56 and the mode in the other class is... (*"this class" refers to the students who did not listen to classical music.*)

72 H: Not the other class the other...

- 73 A: The other people is 60. (*Refers to the students who listened to classical music.*)  
74 H: Try...no not that is the midrange. The mean for the people who did not...  
75 A: 61 and 66. (*The mean for the students who did not listen to classical music is 61 and the mean for the students who did listen to classical music is 66.*)  
76 H: 61 and 66  
77 A: Another fact. They didn't do way better than the other class but they still did better. I mean the other people.  
78 H: Yeah.

Amy and Heather confirmed their earlier claim that students who listened to classical music scored in the sixties and seventies, while the students who did not listen to classical music scored in the forties and fifties. They did this by comparing individual data points and using the statistical tools in TinkerPlots to display the mean, median and mode for both sets of bars. They used a horizontal divider to view the placement of these measures to compare and contrast the test scores of both groups. Amy referred to the measures as “facts” to support their claim. After comparing the calculations for the mean, median, and mode the pair observed that the measures of center were fairly close in value for both groups.

Based on this observation Amy and Heather were reluctant to issue a blanket statement that students who listened to classical music did better on the test. Amy concluded that “They didn't do way better than the other class but they still did better” (line 77). This conclusion was a refinement of their original hypothesis. The session ended with Heather's agreement of Amy's conclusion (line 78).

**Characterization of Pair 6A TinkerPlots session.** Table 2 is a chronological listing of how pair 6A's TinkerPlots session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code.

6A Critical/Creative Thinking	6A Statistical Thinking	Description
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 1	Produces dual circle graphs
Concretize		Discusses good and bad grades
	Extreme Value/Outlier	Refers to lowest score
Causal Relationship		"He must have just been listening to classical music and not studying."
	Spread	"those test scores are really close together"
	Modifying Representation	Modifies dual circle graphs
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 2	Produces multiple circle graphs
	Extreme / Outlier	Uses word "outlier"
	Different Group Size	Notices difference in data set size
	Distribution Chunk	"There are more circles lower."
Hypothesis		"I think it helps."
Verification of Hypothesis	Modifying Representation	Modifies multiple circle graphs
	Classifying	Makes bars red and green
	Sequencing	Orders data from least to greatest
	Representation 3	Produces bar graph
Restatement of Hypothesis		"It shows it helps us."
Causal Relationship Verification of Restatement of Hypothesis	Extreme/Outlier	Comments about highest score "She is just an exceptional student."
	Range	Refers to data in the ranges from 60 - 70 and 40 - 50
	Distribution Chunk	Compares scores in 60s and 70s to scores in the 40s and 50s
	Spread	"big clump of people right there"
	Extreme/Outlier	Comments about low score
Hypothesis with caveat		"it help you score better test, um, it might help you score higher than you usually score, but you should not get so absorbed in the classical music you don't study"
Causal Relationship		"So you shouldn't get so absorbed in the classical music that you don't study."
Verify Hypothesis with caveat	Modifying Representation	Modifies bar graph
	Classifying	Splits data into two columns
	Sequencing	Orders data from least to greatest
	Representation 4	Produces Dual Bar Graphs
	Extreme / Outlier	Comments on a high score
	Range	Refers to range "60 to 78"
	Distribution Chunk	Discusses scores in 60s and 70s
Extreme/Outlier	Comments on low score	

	Average	Finds mean, median and mode
Conclusion		“They didn’t do way better than the other class but they still did better.”

*Table 2. Chronological List of Codes for Pair 6A TinkerPlots Session*

The statistical techniques Amy and Heather used in the creation of their graphs and analysis of the data were classifying, sequencing, extreme/outliers, range, distribution chunks, and recognition of different group sizes. The critical and creative thinking elements that the pair used in the creation of their graphs and analysis of the data were hypothesizing, verification, concretizing, and causal relationships.

The girls’ work on this task can be characterized by an analysis cycle that included analyzing the graphs they created, stating a hypothesis about which group of students did better on the Classical Music test, and attempting to verify the hypothesis. This cycle was repeated three times during the session. What prompted these cycles was the pair being cautious about finalizing their hypothesis. Amy exhibited restraint when she stated “We shouldn’t make any conclusion yet. Let’s put it into a bar graph.” Subsequently, everything the pair did after generating their initial hypothesis was for the purpose of verification.

One characteristic of the pair’s work that was a result of their analysis cycles was the refinement of their original hypothesis. As the pair rotated through the analysis cycle three times, their hypotheses slightly changed with each cycle. In the second cycle they simply restated their initial hypothesis with no changes. In the third analysis cycle they added a caveat to the original hypothesis: “You should not get too absorbed in the music that you don’t study.” And for the conclusion they determined that “They didn’t do way



better than the other class but they still did better.” This conclusion was a refinement of their initial hypothesis. It was made after the pair compared the calculations for the mean, median, and mode and observed that these measures of center were fairly close in value for both groups. Based on this observation Amy and Heather were reluctant to state definitively that said students who listened to classical music did better and it helped improve test scores.

Amy and Heather’s work on this task can also be characterized by their use of modal clumps and distribution chunks. For example, the pair’s third representation was a single bar graph where the colors of the two groups of students were interspersed throughout. Due to the contrasting nature of these colors and the height of the bars, the pair was able to view the placement of “chunks” of red and green bars and discuss the spread of the data. They made several observations regarding the placement of the two groups of students in relation to one another on the graph. The pair noticed that there was a large clump of consecutive red bars (i.e., students who did not listen to classical music) toward the left side of the graph that represented lower test scores. Also, they noticed there was a higher concentration of green bars toward the right side of the graph, that is, students who had listened to classical music had higher test scores.

What the pair was describing was a distribution chunk. The clump of red bars toward the right of the distribution was a small subset of the data that represented the students who did not listen to classical music and had lower test scores. This subset was separated from the other students who did not listen to classical music, as well as the entire group of students, and their scores were analyzed together as a group, rather than as individual scores.

Another characteristic of the Amy's and Heather's work in the session was the identification of causal relationships. All of the instances of causal relationships that were demonstrated by the pair centered on the lowest test score and why this student did not fit the pattern for the students who listened to classical music. The girls tried to explain why Sean, who had listened to classical music, scored lower than any of the other students. They suggested a cause or reason for Sean's poor performance on the test. For example, the pair originally hypothesized that listening to classical music would help improve test scores. In an effort to justify why Sean performed so poorly on the test, they said that "you should not get so absorbed in the classical music you don't study. You should have it low like those people." What they meant was that the people who listened to classical music and made high scores on the test probably had the volume of the music on low so it did not interfere with their concentration while studying. However, Sean must have been listening to the music with the volume such that it distracted him while he studied. Thus, he got "too absorbed in the music." By identifying this causal relationship, Amy and Heather felt that they had adequately explained why Sean, the outlier, did not fit this overall pattern and trend.

In summary, Amy and Heather's work in the TinkerPlots was characterized by repeating the analysis cycle three times, using the mean, median, and mode feature of TinkerPlots to verify their hypothesis, refinement of their original hypothesis, using distribution chunks in their analysis, and the identification of causal relationships.

**Differences in thinking when working in the two mediums.** This section describes the differences in the pair's thinking between the two mediums. In paper and pencil the pair created one graph and in TinkerPlots they created four graphs. In the

analysis of these graphs, a significant difference in Amy and Heather's thinking in TinkerPlots was that they used the context of the data in their analysis. In TinkerPlots the pair made use of concretization and causal relationships. This was not the case in the paper and pencil task as they never attempted to contextualize the data. In TinkerPlots, all of the instances of causal relationships that were demonstrated by the pair centered on the lowest test score (which was also an outlier) and why this student did not fit the pattern for the students who listened to classical music. The girls tried to explain why Sean, who had listened to classical music, scored lower than any of the other students. They suggested a cause or reason for Sean's poor performance on the test. In TinkerPlots, the outlier was easily recognizable and prominent due to the contrasting colors and sequencing of the data from least to greatest. The girls could see the outlier and felt a need to explain it. In contrast, in their paper and pencil representation the outlier was not visible. It had been deleted from the data set when they were modifying their graph in an effort to equalize the group sizes.

The pair's work in both mediums can be characterized by an analysis cycle that included analyzing the graph they created, stating a hypothesis about which group of students did better, and attempting to verify the hypothesis. In paper and pencil this cycle was repeated twice during the session and in TinkerPlots it was repeated three times. In paper and pencil the pair's verification in both cycles of analysis was prompted by Heather's difficulty in dealing with the fact that the data sets did not contain the same number of data points. In TinkerPlots, verification was prompted due to the pair being cautious about finalizing their hypothesis. The pair recognized the difference in group size, and made statements regarding it. However, this awareness did not directly affect

their method of analysis. Their analysis cycles were used to obtain more evidence to verify their hypothesis and inform their conclusion.

In paper and pencil the result of the pair's two analysis cycles was Heather being able to understand that the students who took Driver's Education did better on the test despite having a fewer number of students. In TinkerPlots, the outcome of the analysis cycles was the refinement of their original hypothesis. As the pair rotated through the analysis cycle three times, their hypotheses slightly changed with each cycle.

Amy and Heather used some unconventional techniques in their work in paper and pencil. For example, the "skimming-off" method they employed to modify their back-to-back stemplot was not a well-reasoned approach and yielded an inaccurate representation of the data. Another unconventional technique that the pair used in their paper and pencil session was "approximate averages." The pair used the term "average," but did not calculate the arithmetical mean. What they meant by "average" was that they could see a pattern in the data that the majority of the students in a group scored in a particular range.

In contrast, in TinkerPlots, the pair used the tools available to calculate the mean, median and mode accurately, quickly, and efficiently. They also had no problem modifying and transforming their graphs into accurate representations of the data. TinkerPlots enabled the girls' to create and transform graphical representations that maintained the integrity of the data and also provided them with accurate measures of center. In contrast, when left on their own with only paper and pencil, they relied on informal, unconventional techniques that resulted in a distorted picture of the data.

## Pair 6B

**Summary of paper and pencil session.** Melinda and Tonya began the Driver's Education task (see Appendix C) by splitting the cards into two piles, those that took Drivers' Education and those who did not take Drivers' Education. They ordered the cards in each pile by score from greatest to least. They noticed that there were more people that did not take Drivers' Education and recognized this would be an issue in their analysis of the data (episode 1, line 10). Their initial analysis is presented in Episode 1.

### Episode 1

- 8 T: We could find the average for all of them. Yeah.  
9 M: Oh sorry.  
10 T: The problem is that more people did not take Drivers' Ed.  
11 M: Yeah. Look at it real quick these scores are almost the same as some of these.  
12 T: Well let's do this. What if we like, we put people who have like um, half of 70 is 35. What if we have like over 50% listed...  
13 M: In one column see how many...  
14 T: And then and then um on the... we can't have fractions because... 'cos...  
15 M: There's a lot more people.  
16 T: Yeah.  
17 M: Well let's do that then. Okay. So should we write down?  
18 T: Um I am just going to set these first. Then do that okay. Should we count 35 as above? So that would be at 50%.  
19 M: Yeah because it is (inaudible)...  
20 T: Okay I will leave it there. Should we just look for something under 35? Let's see...  
21 M: There is nothing - everyone went over half so we can't do that.  
22 T: Okay yeah except the Seth kid.  
23 M: But he didn't do...  
24 T: Okay um maybe we should try it the like 75%.  
25 M: Yeah, B, got a B average.

The pair decided to find the number of students in each ordered list that scored above 50% (line 12), then compare the number of students in both groups scoring over this percentage (line 18). They decided to use this type of analysis to account for the difference in data set size (lines 10 - 17). They calculated 50% of the 70-point test was a score of 35 (line 12). Looking through the student scores they found that every student

scored above 50%, except one student who took Drivers' Education – Seth (lines 21 - 23). Due to this, they knew that their plan for comparing the number of students in data sets that scored above 50% would not work (line 22). They modified the test percentage to be used for the comparison to 75% (line 24), noting that this would be a B grade (line 25).

Episode 2 illustrates how Melinda and Tonya calculated 75% of the possible number of points on the test for both data sets.

### Episode 2

- 55 M: Well let's let's give, 'cos 50% is 35 points. We have to find half of 35.  
56 T: Yeah.  
57 M: It would be 17.5.  
58 T: Yeah.  
59 M: Wait wait it would be 34. What is half of 34? 17.5. Put then you have to add 35. Then that would be 75%. So anyone who go 52.5 or up they would get 75%.  
60 T: Mmm yeah so let's just 52.  
61 M: Yeah okay.  
62 T: No 53.
- 63 M: 53 okay, and that is that is pretty much 75%. So I am just going to put 53.  
64 T: So that is to get at least a C. So how many people?

The pair calculated 75% of the 70-point test was a score of 52.5, by taking half of 70 which is 35, then half of 35 is 17.5, then adding 35 and 17.5 for a result of 52.5 (lines 55 - 59). Then they rounded 52.5 to 53 (line 63). They noted a score of 75% on the test would be a C grade – so they would be looking for students who scored “at least a C” on the test (line 64).

They created a chart that was divided into four columns. The first two columns were for the number of students who scored below 75% - one for those who took Drivers' Education and one for those who did not. The second two columns were for the number

of students who scored above 75% - one for those who took Drivers' Education and one for those who did not.

The pair counted the number of students in both groups that scored 53 points and above. They determined that 8 out of 13 students who took Drivers' Education scored above 75%. And 16 out of 35 students who did not take Drivers' Education scored over 75%. Once the number of students that scored above 75% was counted, the table fell by the wayside and the pair did not use it. Instead they wrote the fractions  $8/13$  and  $16/35$  on a sheet of paper separate from the chart. In Episode 3 the girls compare the fractions  $8/13$  and  $16/35$ .

### Episode 3

- 99 M: If we split that in half we get 8 out of 17.5.  
100 T: So we know that this class, more people did better. Not this class, but the people who took it.  
101 M: 8 and 17.5  
102 T: Because they both have 8  
103 M: Oh, they both have...  
104 T: 8 pieces (*Comparing 8/17.5 to 8/13*). If you...  
105 M: So these people have...  
106 T: Right  
107 M: Yeah. But this is out of ... well yeah it would still be this one because this is... uh-huh  
108 T: They both have 8 of these pieces... these pieces smaller...  
109 M: So even though...  
110 T: So even though more people took the other class the people who took Driver's Ed. had higher scores.  
111 M: The people who took it they got...  
112 T: Just let me recount this. I want to make sure.

To compare the two fractions Melinda and Tonya took  $16/35$  and divided both the numerator and denominator by 2 – coming up with  $8/17.5$  (line 99). Now the two fractions had the same numerator ( $8/13$  and  $8/17.5$ ). Once this calculation was made, they immediately hypothesized that the people who took Driver's Education did better, by

reasoning that because the dominator of  $8/13$  was smaller than the denominator of  $8/17.5$ , the 8 pieces of the numerator of  $8/13$  were larger than the 8 pieces of the numerator of  $8/17.5$ , making  $8/13$  the largest fraction (lines 100 - 111). They reiterated their hypothesis, saying that, “So even though more people took the other class the people who took Driver's Ed. had higher scores” (line 110).

The pair exhibited caution and decided to check their work before accepting this hypothesis as their final conclusion (line 112). They recounted the data to check to see that their fractions were constructed correctly. After determining that they were, the pair stated their conclusion “The people who took the class got a better grade.”

**Characterization of Pair 6B paper and pencil session.** Table 3 is a chronological listing of how pair 6A’s paper and pencil session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

6B Critical/Creative Thinking	6B Statistical Thinking	Description
	Classifying	Sorts data cards into two piles
	Sequencing	Orders data cards
	Average	“We could find the average for all of them.”
	Different Group Size	“More people did not take Driver’s Ed.”
	Different Group Size	“There’s a lot more people.”
	Extreme/Outlier	Discusses lowest score.
	Modify	Change percentage for comparison from 50% to 75%
Concretize		“Yeah, B, got a B average.”
Concretize		“So that is to get at least a C.”
	Representation	Constructed fractions $8/13$ and $16/35$
	Modify Representation	Simplified $16/35$ to $8/17.5$
Hypothesis		“So we know that this class, more people did better. Not this class, but the people who took it.”
	Proportional Reasoning	Compares $8/13$ to $8/17.5$
	Different Group Size	“So even though more people took the



Restatement of Hypothesis		other class the people who took Driver's Ed. had higher scores."
Verify		Recounts data to determine fractions constructed correctly
Concretize		Discusses grades
Conclusion		"The people who took the class got a better grade."

*Table 3.* Chronological List of Codes for Pair 6B Paper and Pencil Session

In the first part of the session Melinda and Tonya primarily used statistical thinking elements to create their representation and analyze the data. Once their representation was completed, the focus shifted to critical and creative thinking elements. The statistical techniques the pair used during the session were classifying, sequencing, average, extreme/outlier, recognition of different group size, proportional reasoning, representation, and modification of representation. The critical and creative thinking techniques the pair used were concretizing, hypothesizing, and verification.

Melinda and Tonya's work on this task can be characterized by the technique they used to create and compare the fractions they constructed. They employed a method which I called the "cut-off" technique. The pair decided to use a score of 75% to be the cut-off between "good" and "bad" scores. Above 75% was considered "good" and a score below 75% was considered "bad." The number of students who scored above the cut-off were counted and included in the numerator of the fraction for their respective group. Then these counts were placed over the total number of students in the group. The pair chose this method of analysis to account for the difference in size of the data sets. The fractions made it possible to compare the data sets on an equal basis; there were no longer different group sizes to deal with.

The choice of 75% as a cut-off can be characterized by the pair's recognition of the extreme low score. The pair originally attempted to create two fractions, one for each group of students that represented the proportion of students that scored above 50% on the test. However, they observed that out of all of the students, only one scored below 50%. Therefore, they decided to compare the proportion of students from both groups that scored above 75% on the test.

The girls' work on this task can be characterized by an analysis cycle that included analyzing the graph they created, stating a hypothesis about which group of students did better on the Driver's Education test, and attempting to verify the hypothesis. The cycle was repeated once during the session. Creating the fractions was a very time intensive process, leaving little time for analysis. The pair did not generate their hypothesis until after the fractions were constructed near the end of the session. The pair's verification of their hypothesis can be characterized as weak. Rather than attempting to validate their hypothesis, the pair merely checked the accuracy of their fractions.

Another characteristic of Melinda and Tonya's work during the session was adding context to the data. Concretizing is making a general idea specific by giving examples and applications which will make an abstraction meaningful. The pair demonstrated three instances of concretizing during the session, taking student test scores and mentally converting them to grades. For example, they made statements such as "Yeah, B, got a B average," and "So that is to get at least a C." Since the pair was familiar with grades in their everyday school experience, this put the students test scores into a context they could better understand.

In summary, Melinda and Tonya's work in the paper and pencil environment was characterized by the "cut-off" technique they used to create their representation, completing the analysis cycle once during the session, and adding context to the data.

**Summary of Pair 6B TinkerPlots session.** The second task Melinda and Tonya completed required them to determine whether students studying for a mathematics test while listening to classical music performed better than those who did not (see Appendix B). The data in TinkerPlots were presented to the participants in a random arrangement of 48 circles in differing shades of blue. Each circle represented a student and his or her score on the mathematics test. Each circle was filled with a shade of blue that reflected the student's test score. Darker shades of blue represented higher scores whereas lighter shades of blue represented lower scores. Refer to Appendix A for an image of the way the data were presented to the students.

This session began with the pair reviewing the printed version of the question. Melinda asked me for the total possible points on the test and upon hearing that it was 80 points, she said "Nobody got a perfect" then a few seconds later she said "except Laura." At that point the pair decided that Tonya would be in charge of operating the mouse in TinkerPlots. The first option the pair took was to create a bar graph that was ordered by test score and colored to show whether or not the students listened to classical music (Figure 11). The students that listened to classical music were represented with the color green and the students that did not listen to classical music were represented with the color red.

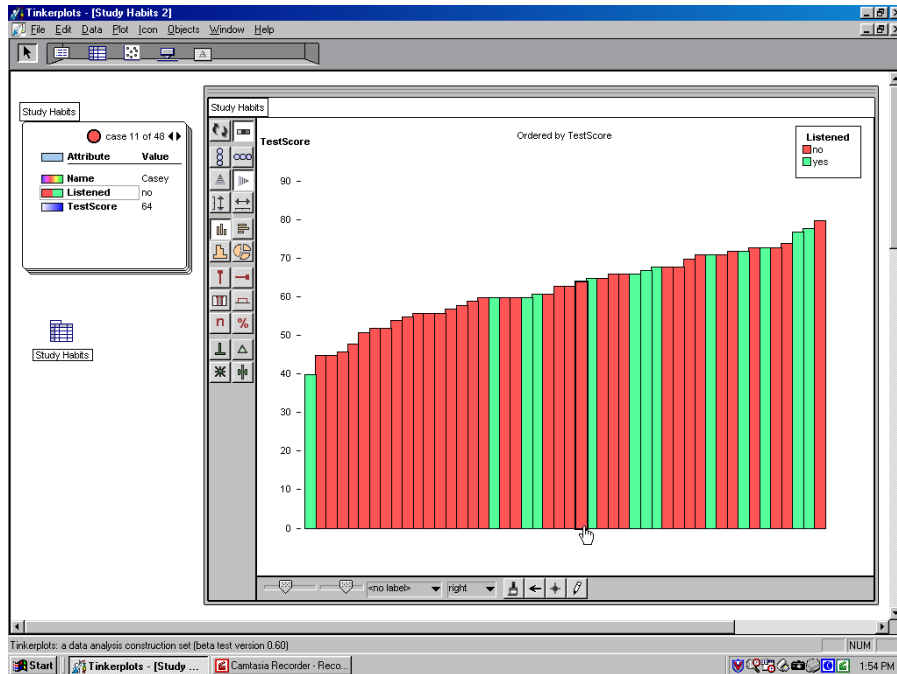


Figure 11. Bar Graph

The pair manipulated Figure 11 and created several other graphs quickly, but did not analyze them. They recreated Figure 11 and without any discussion stated: “They didn’t do badly and they listened” (episode 1, line 48). Episode 1 illustrates how Melinda and Tonya analyzed the bar graph they constructed.

#### Episode 1

- 48 M: They didn't do badly and they listened.  
 49 T: Um, okay, who was that person? Could we check and see who that is? The underachiever.  
 50 M: Sean.  
 51 T: So they some people listened and still got bad grades. But the people who listened did get good grades too.

Melinda and Tonya explored the graph by clicking on the bars to determine student test scores. They found that the shortest bar represented the student who scored the lowest overall – Sean who listened to classical music (lines 49 - 50). Sean was referred to as “the underachiever” (line 49). Tonya noticed that not all of the green bars were high scores,

but some of them were very high. Regarding this, Tonya said “So they some people listened and still got bad grades. But the people who listened did get good grades too” (line 51). This statement was similar to line 48. Both statements were analytical in nature, helping the pair compare the data sets, but not answering the question being asked.

The pair altered Figure 11 by separating it into two side-by-side bar graphs (Figure 12). The bar graph on the left represented students that listened to classical music and the graph on the right represented students that did not listen to classical music.

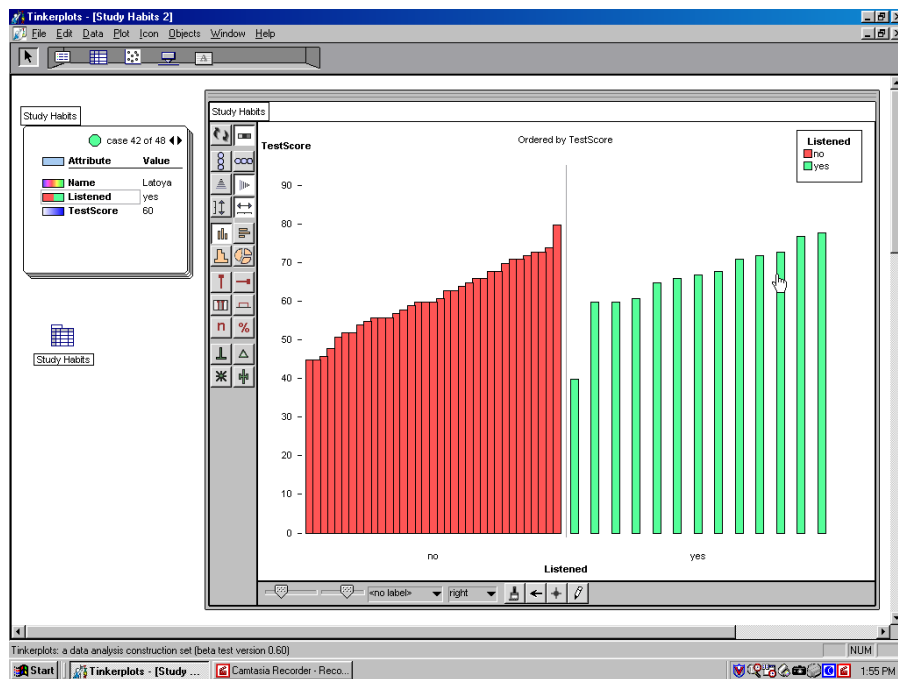


Figure 12. Side-by-Side Bar Graph

Episode 2 illustrates how Melinda and Tonya analyzed the side-by-side bar graph they created.

### Episode 2

- 64 T: If you compare them you can't really because their...
- 65 M: Cos'...
- 66 T: ...not equal.
- 67 M: Lots of people who did listen to classical music did good pretty good test scores accept this person.

In their analysis of figure 12, the pair recognized that the number of students who took driver's education was different than the people who did not take driver's education (lines 64 - 66) and that comparing the two data sets would be difficult due to this (lines 64 - 66). Melinda was able to see that most of the green bars represented high scores, except for one, which was the lowest score of both of the two groups (line 67).

Melinda and Tonya continued their analysis of the data by creating a different type of graph; a multiple circle graph consisting of eight circles. The circles were stacked vertically and organized into bins with a range of 5 points. Each circle had sections that represented an individual test score. The green sections of each circle represented the students who listened to classical music and the red sections red represented the students that did not listen to classical music.

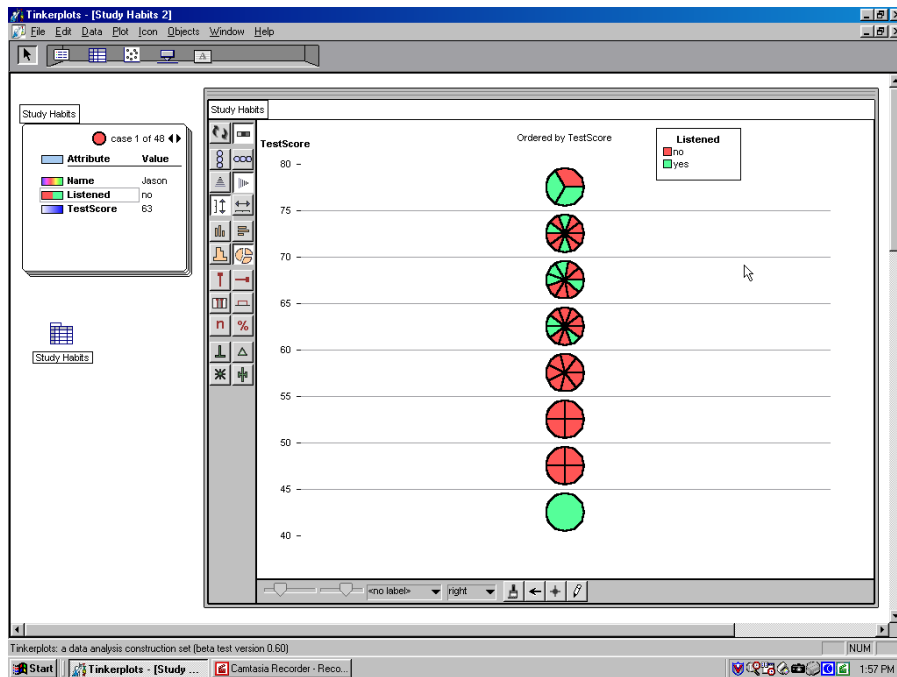


Figure 13. Stacked Multiple Circle Graph

Episode 3 illustrates how Melinda and Tonya analyzed the stacked multiple circle graph

they created.

### Episode 3

97 M: So at the very top circle two out of three people listened.

98 T: yeah

99 M: Who got above 75.

100 T: apparently

101 M: Order this, there we go. Okay that that would be easier.

102 T: Like so a lot of the people who didn't listen

103 M: The greens are towards the top actually

104 T: Yeah, except for this...

105 M: But that's only one person

106 T: Yeah, that's an under-achiever.

107 M: Um, because the reds are here and most of the greens are up here.

Melinda and Tonya observed that the circle on indicated that 2 out of 3 students listened

to classical music and these students scored above 75 (lines 97 - 100). It was also

observed that most of the greens were toward the top of the graph and more reds toward

the bottom of the graph (lines 103, 107), except for the green circle on the bottom which

represented the test score of Sean – the student who scored the lowest (lines 104 - 106).

The pair continued their analysis of Figure 13 in Episode 4.

### Episode 4

113 M: So I think it's true that listening to classical music can help.

114 T: Because everyone got...

115 M: For some people...

116 T: ...above 60.

117 M: Except this person.

118 T: Well that person probably did not...

119 M: Sean

120 T: okay... so... yeah

121 M: And so I think that...

122 T: "Classical music probably did help."

123 M: Some most people... it kinda depends on what their brains like I guess.

124 M: Maybe they just...

125 T: Listened but they didn't do anything.

126 M: Yeah they didn't study, so...

127 T: But these people did.

128 M: We think that classical music can help you study. Maybe it relaxes the mind.

The pair hypothesized "So I think it's true that listening to classical music can help" (line

113). They came to this hypothesis since everybody who listened to classical music scored above 60 points on the test; with the exclusion of the lowest overall score (lines 114 - 119). The observation that the student with the lowest score overall listened to classical music, yet did not do well like the rest of the students who listened to classical music, prompted the pair to refine their hypothesis and conclude that classical music “probably” did help, but “it kinda depends on what your brains like I guess” (lines 122 - 123). They explained that some people may have listened to classical music but did not study (lines 124 - 127). This statement was made in direct reference to Sean, the student with the lowest score, as a reason why he scored so low, yet listened to classical music. The session ended with Melinda’s conclusion that classical music can help you study because “it relaxes the mind” (line 128).

**Characterization of Pair 6B TinkerPlots session.** Table 4 is a chronological listing of how pair 6B’s TinkerPlots session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, in addition to what is evident in the episodes presented.

6B Critical/Creative Thinking	6B Statistical Thinking	Description
	Extreme/Outlier	“Nobody got a perfect, oh except Laura.”
	Sequence	Orders data from least to greatest
	Classify	Splits data into two columns
	Representation 1	Produces bar graph
	Modifying Representation	Modifies bar graph
Analysis-Based Observation		“They didn’t do badly and they listened.”
	Extreme/Outlier	Refers to lowest score
Causal Relationship		“The underachiever. Sean.”
Concretize		“So the people who listened still got bad grades. But the people who listened did get good grades too.”
Analysis Based Observation		
Knowledge Building	Modifying Representation	Modifies bar graph.



	Sequence	Orders data from least to greatest	
	Classify	Splits data into two columns	
	Representation 2	Produces side-by-side bar graph	
	Different Group Size	Notices difference in size of data sets	
Analysis Based Observation		“Lots of people who did listen to classical music did good, pretty good test scores accept this person.”	
Knowledge Building	Modifying Representation	Modifies side-by-side bar graph	
	Sequence	Orders data from least to greatest	
	Classify	Splits data into two columns	
	Representation 3	Produces stacked multiple circle graph	
	Concretize	Discusses good grades	
	Proportional Reasoning	“So at the very top circle two out of three people listened.”	
	Distribution Chunk	“The greens are towards the top actually.”	
	Causal Relationship	Extreme/Outlier	Discussion low score
			“Yeah, that’s an underachiever.”
Distribution Chunk		“because the reds are here and most of the greens are up here”	
Hypothesis		“So I think it’s true that listening to classical music can help.”	
	Extreme/Outlier	Comments on low score	
Re-statement of hypothesis		“Classical music probably did help.”	
Causal Relationship		“it kinda depends on their brains I guess”	
Causal Relationship		“listened but they didn’t do anything”	
Conclusion		“we think that classical music can help you study”	
Causal Relationship		“Maybe it relaxes the mind.”	

*Table 4. Chronological List of Codes for Pair 6B TinkerPlots Session*

The statistical techniques that Melinda and Tonya used in the creation of their graphs and analysis of the data were classifying, sequencing, modification, extreme/outlier, proportional reasoning, distribution chunk, and recognition of different group sizes. The critical and creative thinking elements that the pair used in the creation of their graphs and analysis of the data were analysis based observations, knowledge building, concretizing, and causal relationships.

The critical and creative thinking of this pair can be best characterized by their cycle

of analysis. Rather than hypothesizing, the pair used a variation of this thinking element, an “Analysis-Based Observation.” I first identified this new thinking element while analyzing this pair, and subsequently found it demonstrated by other pairs. They made statements that were not direct answers to the question asked. These statements did not meet the criteria for a hypothesis. The definition of a hypothesis is that students will develop a testable explanation for a given situation. However, these statements were valuable to the students understanding of the data they were analyzing. They were analytical in nature but did not present a testable explanation. Each of the analysis-based observations were about the students who listened to music doing “good” but it is never stated that the students who listened to classical music did better than the students who did not listen to classical music. Thus, the questions that were asked were never answered.

The pair never technically verified their analysis-based observations. After an analysis-based observation they continued to explore and analyze the data, in an effort to learn more while keeping an open mind. They were not trying to prove that classical music helped the students’ test scores. I first identified this variation of the thinking element verifying when analyzing this pair, and subsequently found it demonstrated by other pairs. The term I used for this type of analysis was “Knowledge Building.” When the pair used knowledge building, in lieu of verifying, they continued analysis, without trying to prove if an answer to the question that they had generated was true. Instead, they just continued to analyze the data, or “build knowledge about it.”

Throughout the session the pair’s statistical thinking can be characterized by their use of distribution chunks and proportional reasoning. In their analysis of Figure 13 they

used both of these statistical thinking elements. They proportionally reasoned that two out of three students who listened to classical music score above 75%. They made a visual inspection of the data and observed that the green sections were up higher on the graph and represented higher scores than the red sections – the students who did not listen to classical music. This qualitative comparison that the pair was describing was a distribution chunk. They recognized that despite there being fewer students who listened to classical music, most of their scores were up higher on the graph – meaning that they scored higher than the students who did not listen to classical music. It should be noted that the pair mentioned the difference in group size once during the session. However, they never overtly acted on this difference, but were able to appropriately compare the two groups using distribution chunks.

Another characteristic of Melinda and Tonya’s work in the session was the identification of causal relationships. The pair’s identification of causal relationships were frequently paired with the extreme low score. Melinda and Tonya suggested reasons why Sean scored so low, yet had listened to classical music.

One characteristic of the pair’s work that was a result of their analysis cycles was the refinement of their original hypothesis. The pair hypothesized: “So I think it's true that listening to classical music can help.” After stating this, the pair considered the effect of the extreme low score. The observation that the student who scored the lowest overall listened to classical music, yet did not score as well as the rest of the students who listened to classical music, prompted the pair to modify their original hypothesis and conclude that classical music “probably” did help and “it kinda depends on your brains like I guess.” They explained that some people may have listened to classical music but

did not study. This statement was made in direct reference to Sean, the lowest score, as a reason why he scored so low, yet listened to classical music. They continued to reason that classical music can help you study because “it relaxes the mind,” qualifying their hypothesis by suggesting causes for why listening to classical music helped improve test scores.

In summary, Melinda and Tonya’s work in the TinkerPlots was characterized by their use of analysis-based observations, knowledge building, proportional reasoning, using distribution chunks in their analysis, the recognition of different group sizes, the identification of causal relationships, and the refinement of their hypothesis.

**Differences in thinking when working in the two mediums.** This section describes the differences in the pair’s thinking between the two mediums. A major difference in their work during the sessions was how they represented the data. In paper and pencil the pair represented the data with two fractions. In TinkerPlots they created four different graphs.

A significant difference in Melinda and Tonya’s thinking was that the pair identified causal relationships in TinkerPlots. This was not the case in paper and pencil task as they never demonstrated the use of a causal relationship. The pair’s identification of causal relationships in TinkerPlots were frequently paired with the extreme low score. These statements rationalized how Sean, a listener of classical music, scored so low. Causal relationships also helped them rationalize their hypothesis that the students who listened to classical music scored higher overall.

In TinkerPlots the outlier was highly visible in all the graphs the pair created. The girls could see the outlier and felt a need to explain it. In contrast, in paper and pencil the

pair represented the data as two fractions. The outlier, as well as the rest of the individual scores, were not visible as they had been combined into aggregates of the data set.

In both mediums, the pair demonstrated analysis cycles involving analyzing the representation they created, stating a hypothesis about which group of students did better, and attempting to verify the hypothesis. There were differences in the cycles between the environments. Creating the fractions in paper and pencil was a very time intensive process, leaving little time for analysis. The pair did not generate their hypothesis until after the fractions were constructed near the end of the session and for verification they just checked their accuracy. Their analysis cycle can be characterized as weak; their verification was more of an afterthought as it was not used to obtain more evidence to support their conclusion.

In TinkerPlots the pair's analysis was a lengthy process and they cycled through it three times. During these cycles the pair used analysis-based observations and knowledge building, in lieu of hypothesizing and verifying, to obtain more evidence to verify their hypothesis and inform their conclusion. These cycles were aided by the pair's ability to modify existing graphs and create new ones quickly and efficiently. In TinkerPlots the subjects were freed from the tedium of mathematical calculation that they endured in the paper and pencil task. TinkerPlots allowed them to rapidly and easily create graphs that were complete, representative, and appropriate for the data set. These analysis cycles resulted in the refinement of their original hypothesis to include causes for why listening to classical music helped improve test scores. These refinements were prompted by the consideration of the effect of the extreme low score.

There was a difference in the pair's method of analysis in the two environments. In

paper and pencil the pair quantified the data, lumping all the data into two fractions and then comparing their size. In TinkerPlots the pair visually inspected the data making qualitative comparisons of distribution chunks they had identified. These methods enabled the pair to compare the data sets despite their difference in group size. In paper and pencil, the fractions equalized the group size which made it possible to compare the data sets on an equal basis. In TinkerPlots, the pair mentioned the difference in group size during the session, however, they never overtly acted on this difference. Melinda and Tonya were able to appropriately compare the two groups using distribution chunks.

### **Pair 6C**

**Summary of paper and pencil session.** Brad and Joey began the Driver's Education task (see Appendix C) session by visually searching for the high and low scores in both data sets. Their initial discussion is presented in Episode 1.

#### Episode 1

- 1 J: To begin with...
- 2 B: The lowest Drivers' Ed. score is 30.
- 3 J: 30, alright so...
- 4 B: And the highest Drivers' Ed. score is 68. 30 to 68.
- 5 J: Alright.
- 6 B: 30 to 68 and for without it's...
- 7 J: Alright so the lowest was, we said 30 with Drivers' Ed. The highest that took Drivers' Ed. was, what did you say? 67?
- 8 B: 60
- 9 J: 67
- 10 B: Yeah 68, 68, Peter.
- 11 J: Uh, yeah, yeah.
- 12 B: 68 and the low, this one its 35.
- 13 J: No 38.
- 14 B: Yeah it's 38, my bad, my bad.
- 15 J: And the highest is...
- 16 B: No it's 35. You messed up Joey. And the highest is...
- 17 J: 55... 63. No, no Luke is 64.
- 18 B: Nope Lindsay.
- 19 J: Yeah 70.

The high and low test scores were found to be 30 and 68 for the students who took Drivers' Education and 35 and 70 for students who did not take Drivers' Education. Once the pair had identified these scores they recorded them on paper (Figure 14).

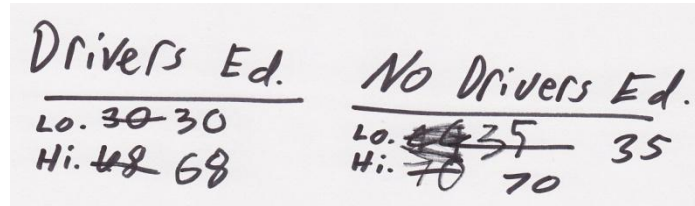


Figure 14. High and Low Scores

After Episode 1, without further analysis, the pair hypothesized that “Drivers Ed. did not help very much” (episode 2, line 28). In Episode 2, Brad and Joey interpreted the high and low scores they identified.

#### Episode 2

- 28 B: Drivers' Ed. did not very much. Did not help a whole lot.
- 29 J: Right
- 30 B: Because the people scored kind of bad. (*Referring to the students who took Drivers' Education.*)
- 31 J: They scored a lot lower, well not, well only...
- 32 B: Yea they scored lower than the... (*Brad points from the students who took Drivers' Education to the students who did not take Drivers' Education.*)
- 33 J: And they scored...
- 34 B: Maybe this Seth guy. (*Referring to the student who had the lowest score of those who Drivers' Education.*)
- 35 J: They got lower in general. They got lower in general.
- 36 B: They got lower.
- 37 J: So it did not really help.
- 38 B: It did not really help.
- 39 J: That is our conclusion for number one it did not help.
- 40 B: Nope
- 41 J: Nope
- 42 B: Diddly-squat. Number two, what kind of conclusion can you make about...
- 43 J: Drivers' Education just takes up your time. It does not help you very much.
- 44 B: Not really that much. That is our conclusion.

The pair reasoned that Driver's Education did not help very much (line 28) since the students who took drivers education “They got lower in general” (line 35). What they

meant by this was that lowest score for the students who took Drivers' Education was lower than the lowest score for the students who did not take Drivers' Education.

The pair's discussion from line 29 through 36 was an attempt to validate their hypothesis. They discussed the extreme low score and commented that the students who took Drivers' Education "scored kind of bad" (line 30). The pair continued the session by re-stating their hypothesis several times (lines 37 – 39) and finally concluded that "It [Driver's Education] does not help you very much" (line 43). Brad provided a reason why Driver's Education did not help improve test scores. As his final statement in the session, he added that "Drivers education just takes up your time" (line 43).

**Characterization of Pair 6C paper and pencil session.** Table 5 is a chronological listing of how pair 6C's session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

6C Critical/Creative Thinking	6C Statistical Thinking	Description
	Extreme/Outlier	Searches for highest and lowest scores.
	Range	Refers to range 30 to 68.
	Extreme/Outlier	Discusses high and low scores.
Hypothesis		"Driver's Ed. did not help very much."
Verification		
		"Because the people score kind of bad."
Concretize	Extreme/Outlier	Discusses lowest score.
Re-statement of Hypothesis		"Driver's Education just takes up your time."
Causal Relationship		
Conclusion		Re-statement of hypothesis. "It does not help you very much."

*Table 5.* Chronological List of Codes for Pair 6C Paper and Pencil Session

The statistical thinking elements that Brad and Joey used in their analysis of the



data were extremes/outliers and range. The critical and creative thinking elements that the pair used were hypothesizing, concretizing, and the identification of causal relationships.

One characteristic of the pair's work is that they did not take all the data into account. They only recorded the high and low test scores for both groups of students. While there were no inaccuracies in Brad and Joey's determination of the extreme scores, they were not representative of the entire data set. None of the values in the middle of the distribution were represented. They only visually inspected the data and based their argument solely on extreme features of the data, ignoring any patterns and trends in the data.

Another characteristic of Brad and Joey's work was an analysis cycle that included a hypothesis that Driver's Education did not affect test scores, attempting to justify their hypothesis, and then re-stating the hypothesis as their conclusion. The pair based this analysis on the extreme low score. They stated that the students took Driver's Education scored lower in general, yet based this off of Figure 14 which only showed the high and low scores for both groups of students. They characterized all the students who took Driver's Education based on the extreme low score.

Brad and Joey's work on this task can also be characterized by their personalization of the data. They demonstrated concretizing and the identification of causal relationships, both elements of critical and creative thinking. Their instances of adding context to the scores were not related to one another – one was regarding the scores of Drivers' Education students being "bad," and other explaining why Driver's Education did not affect test scores.

In summary, Brad and Joey's work in paper and pencil was characterized by their

absence of a graphical representation, the faulty logic their hypothesis was based on, and their personalization of the data.

**Summary of Pair 6C TinkerPlots session.** The second task Brad and Joey completed required them to determine whether students studying for a mathematics test while listening to classical music performed better than those who did not (see Appendix B). The data in TinkerPlots were presented to the participants in a random arrangement of 48 circles in differing shades of blue. Each circle represented a student and his or her score on the mathematics test. Each circle was filled with a shade of blue that reflected the student's test score. Darker shades of blue represented higher scores whereas lighter shades of blue represented lower scores. Refer to Appendix A for an image of the way the data were presented to the students.

Brad and Joey began their work with TinkerPlots by manipulating the default view of the data so that the circles that represented those that listened to classical music while they studied, and those that did not, were arranged in two bins. The pair then ordered the data by test score by clicking on the word "TestScore" in the data card. On the side-by-side stacked dot plot they created (Figure 15), the vertical ruler on the left-hand side of the graph represented the number of test scores in each bin. To view the test score of an individual student, the pair clicked on a circle and the score was displayed in the data card.

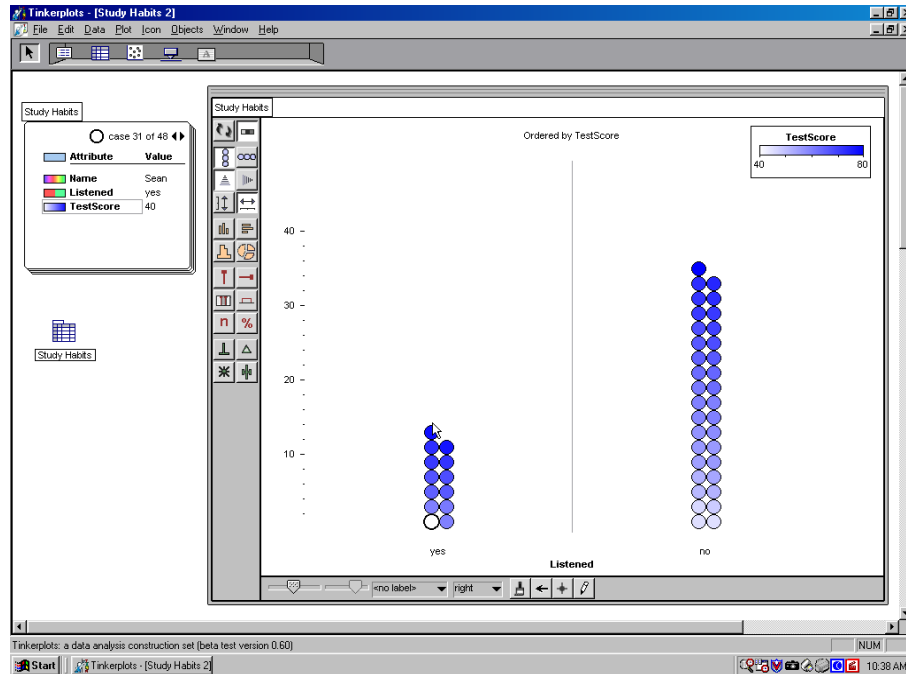


Figure 15. Side-by-Side Stacked Dot Plot

Episode 1 illustrates how Brad and Joey analyzed the side-by-side stacked dot plot they constructed.

Episode 1

- 10 J: Ok, um, so highest test score...
- 11 B: Well this person they got 40 they got an F. *(Referring to the lowest score for the students that listened to classical music.)*
- 12 J: And they listened to classical music, that's not good. No that's 50 percent.
- 13 B: That's an F!
- 14 J: Alright, now the best person got a 78. That is 78 out of 80 so that is very good. *(Referring to the best score for students who took listened to classical music.)*
- 15 B: That is good. Now look at the best person on this side. *(Referring to the students that did not listen to classical music.)*
- 16 J: Now the best person on this one...
- 17 B: Got an 80.
- 18 J: Whoa 80.
- 19 B: Whoa.
- 20 J: Whoa. Ok, it might have helped to have listened to classical music, but if you are naturally smart like me, they will do very well.
- 21 B: Wally
- 22 J: But the person who did the worst here is 45. So the worst person, who did not listen to the music, got even better than the person who did listen to the

classical music and got a 40. And this person who got a test score without listening to classical music like what I just picked got a 100 percent, more than the 78 of the yes...

The pair's discussion and analysis of Figure 15 focused on identifying then comparing the highest and lowest test scores for both data sets. Student's test scores were found by clicking on individual circles and then viewing their information on the data card in the upper left-hand corner of the screen. The high and low test scores were found to be 40 and 78 for the students who had listened to classical music and 45 and 80 for students who had not listened to classical music (lines 2 - 18).

The pair found that the lowest overall test score belonged to a student named Sean who had listened to classical music (lines 11 - 13). They converted his test score of 40 out of 80 points to 50% and stated that this was an F - which was not good.

At the end of Episode 1, Joey compared the lowest scores from both data sets to one another. He stated that the lowest score for the students who had not listened to classical music was 45, which was "even better" than the lowest score for the students who had listened to classical music (line 22). The same comparison was made for the highest scores - a student who had not listened to classical music scored a 100%, which was higher than the best score for the students who had listened to classical music (line 22).

The pair altered Figure 15 by merging the circles representing individual scores into one large circle for each data set. The result was two side-by-side circle graphs (Figure 16). The circle on the left represented students who had listened to classical music and the circle on the right represented students who had not listened to classical music. Both graphs were ordered by test score clockwise from least to greatest starting at

the 3:00 position.

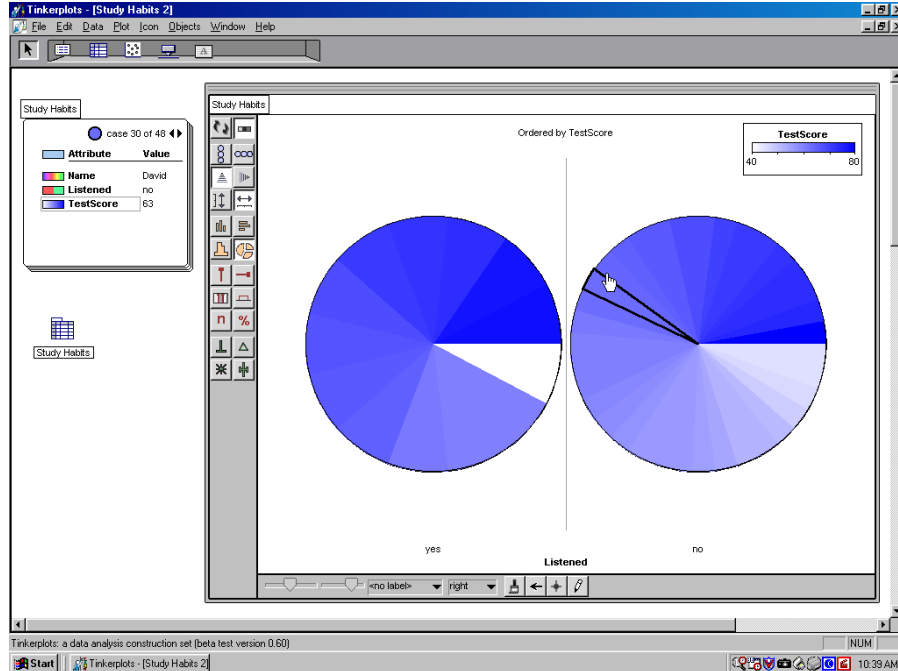


Figure 16. Side-by-Side Circle Graphs

Episode 2 illustrates how Brad and Joey analyzed the side-by-side circle graphs they created.

### Episode 2

- 25 B: This is students that did, that they studied yes. (*Referring to the right circle graph.*)
- 26 J: Well you can...
- 27 B: There is a lot of blues. (*Referring to left circle graph.*)
- 28 J: You can already tell...
- 29 B: And there is a few purples.
- 30 J: You can already tell about more people listened, did not listen because there are a lot more triangles, like these triangles are really small, but compare that to this, this is really big, so in other words there are a lot more people that did not do it. These are like really big. (*The triangles that Joey is referring to are the sectors of the circle graphs.*) (*Comparing both graphs.*)

Brad and Joey observed different aspects of the graph. For example, Brad focused on the color of the graphs and noted that “there is a lot of blues” and “there is a few purples” (lines 27, 29). In the graph on the left-hand side, Brad referred to approximately

three-quarters of it (7:00 to 3:00) as blue, and the other one-fourth (3:00 to 7:00) as purple. There was a “cut-off” at approximately 7:00 where on one side the graph was blue and on the other side the graph was purple.

Joey recognized that there were more segments on the graph that represented those that had not listened to classical music and fewer segments on the graph representing students who had listened to classical music (line 30). Another item he made note of was that the segments of the graph that represented students who has listened to classical music were larger than the segments in the other graph. Based on these observations he concluded that there were fewer students who had listened to classical music than did not listen to classical music.

Brad and Joey continued their analysis of the data by creating a side-by-side vertical dot plot. The graph was separated into two columns along the horizontal axis with test scores on the vertical axis. The left column displayed students that had listened to classical music and the right column displayed those students that had not listened to classical. The vertical axis was segmented into sections of test scores in increments (bins) of seven points each. Within each bin, student test scores, represented by circles, were arranged in horizontal rows, ordered by test score.

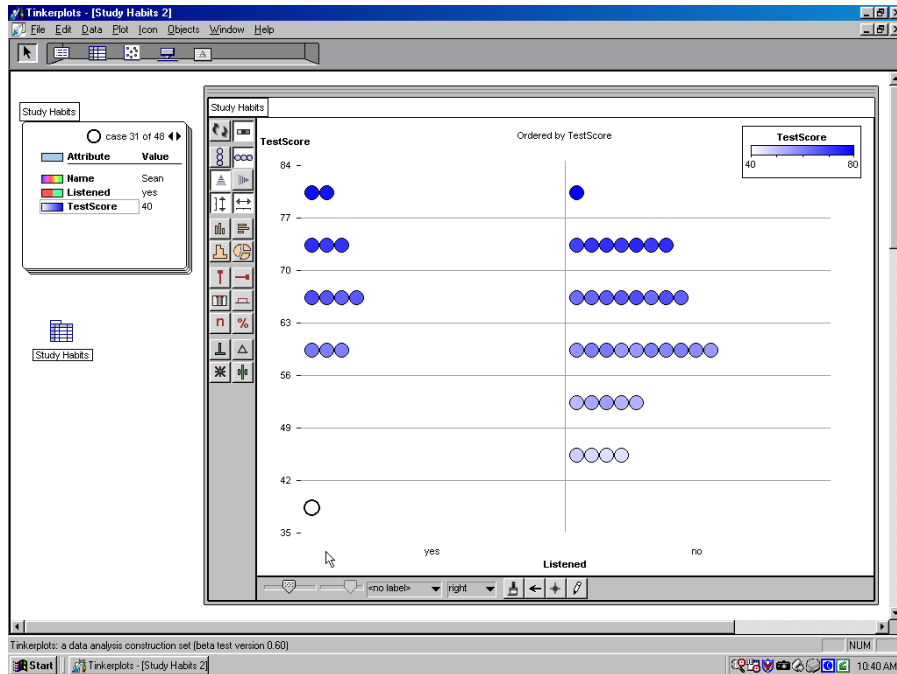


Figure 17. Side-by-Side Vertical Dot Plot

Episode 3 illustrates how Brad and Joey analyzed the side-by-side vertical dot plot they constructed.

Episode 3

- 41 B: This person is plain dumb (*Pointing to the lowest score for the students who listened to classical music.*)
- 42 J: Like Brad.
- 43 B: No like you. So we're, we're going to exclude this dude.
- 44 J: No we're not.
- 45 B: He's an outlier Joey.
- 46 J: Oh, oh he's an outlier.
- 47 B: Yes he is an outlier.
- 48 J: For the people who did listen to it. What is his name, it does not say.
- 49 B: What! This one is locked. (*Brad wanted to exclude the outlier. He randomly clicked on a control at the bottom of the screen. A box opened that said the collection was locked for editing.*)
- 50 J: The other one wasn't locked. (*It is not clear what Joey is referring to, as they had not tried to edit the data before*). Um, this it doesn't say. Sean, Sean apparently did very bad. (*The student with the lowest score of those who listened to classical music.*)
- 51 B: Sean did not.
- 51 J: Well the person that got the best.
- 53 B: Sean lied about listening to classical music.

- 54 J: Toni, wait no Toni.  
55 B: Toni. Toni did bad.  
56 J: Where is the 80?  
57 B: Look for it.  
58 J: Laura got very good for not listening. (*Laura did not listen to classical music. Laura scored the highest score of all students, an 80.*)  
59 B: Laura she studied.  
60 J: She probably listened to pop.  
61 B: She probably listened to rap or something good.  
62 J: But this person Phil probably listened popular and...  
63 B: This guy probably listened to some Beethoven or something, he got an F. (*Referring to Sean.*)  
64 J: Now the first question. How did the test scores... um, it did not really effect.  
65 B: Are you sure?  
66 J: Well look.  
67 B: Maybe only smart people listened to classical music.  
68 J: Like me.  
69 B: Yeah.  
70 J: I prefer ... ok um but I would say would not affect it. That is my hypothesis.  
71 B: Why is that so?

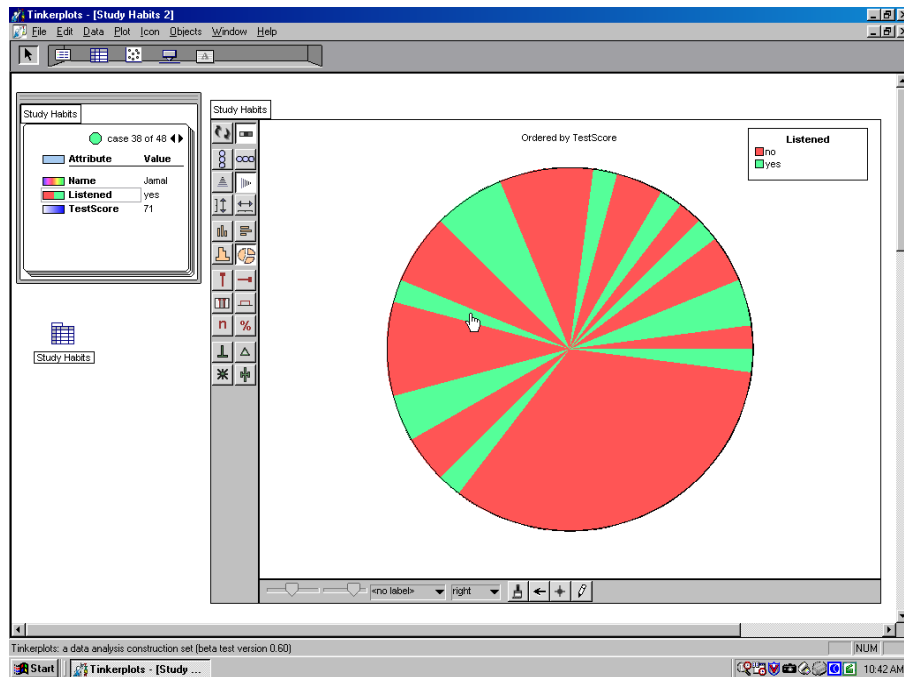
The strategy that Brad and Joey used to analyze Figure 17 was a comparison of individual test scores and focused on the highest and lowest scores. In the process, they identified Sean, who had listened to classical music but had the lowest overall test score (lines 41 - 47). In an effort to explain how this could have happened, the pair suggested a number of possibilities that could have contributed to this phenomenon. To explain Sean's low score, Brad stated that "Sean lied about listening to classical music" (line 53).

Earlier in the session the pair observed that Laura, a student who had the highest overall score, had listened to classical music. In Episode 3, Brad stated "Laura, she studied" (line 58). The pair explained the effect of listening to a certain type of music might have on a student's score. Joey said, Laura "she probably listened to pop," and Brad says "She probably listened to rap or something good" (lines 60 – 61). As for Sean, Brad stated that he got an F because he "probably listened to some Beethoven" (line 63).



Joey provided the first response to the solution of the task (line 64), hypothesizing that listening to classical music did not affect test scores. Brad questioned Joey's hypothesis asking, "Are you sure? Why is that so?"(lines 65, 71). Joey was unable to respond to Brad's question.

In an effort to justify his hypothesis, Joey modified Figure 17 into a circle graph that was ordered by test score and overlaid with whether or not a student listened to classical music (Figure 18). The graph had red and green segments where red represented students who had not listen to classical music and the green represented those that did listen to classical music. The graph was ordered from highest to lowest score with the highest score at the 3:00 position and then the scores decreased in a counterclockwise direction.



*Figure 18. Red and Green Circle Graph*

Episode 4 illustrates how Brad and Joey analyzed the circle graph they created.

#### Episode 4

- 87 B: Yes green is no, yes, red is no, yes.  
88 J: But we don't know their test score  
89 B: This is the highest person and then it goes that way. *(Brad clicked on the red sector at the 3:00 position on the graph. He saw that this sector represented Laura who scored 80. He knew from analysis earlier in the session that this was the highest score for both groups of students.)*  
90 J: But we don't know his test score  
91 B: Does it care? This is the highest person. There's more greens up here. Except for this guy he's outlier.  
92 J: Ok um the red.  
93 B: So there are more greens up here and more reds down here.  
94 J: Wait no.  
95 B: Yes these are all the people with like...  
96 J: Which guy is the high score?  
97 B: This guy is the high score, so then it goes that way. *(Brad was referring to the red sector in the 3:00 position on the graph which represented Laura who had scored the highest of both groups.)*

Brad understood how Figure 18 was organized. He knew where the highest score and outlier (line 91) were placed on the graph and how the scores were ordered from greatest to least. On the other hand, Joey did not understand how Figure 18 was organized. He could not locate the highest score on the graph "Which guy is the high score?" (line 96) and was confused whether or not a student listened to classical music with gender. Brad attempted to help Joey understand the graph; however, Joey revised the graph before the pair had a chance to do more analysis.

The pair altered Figure 18 by splitting the circle graph into two side-by-side circle graphs (Figure 19). The circle graph on the left represented students that listened to classical music and the circle graph on the right represented those that had not listen. Both graphs were ordered by test score clockwise from least to greatest starting at the 3:00 position. Figure 19 was identical to Figure 16.

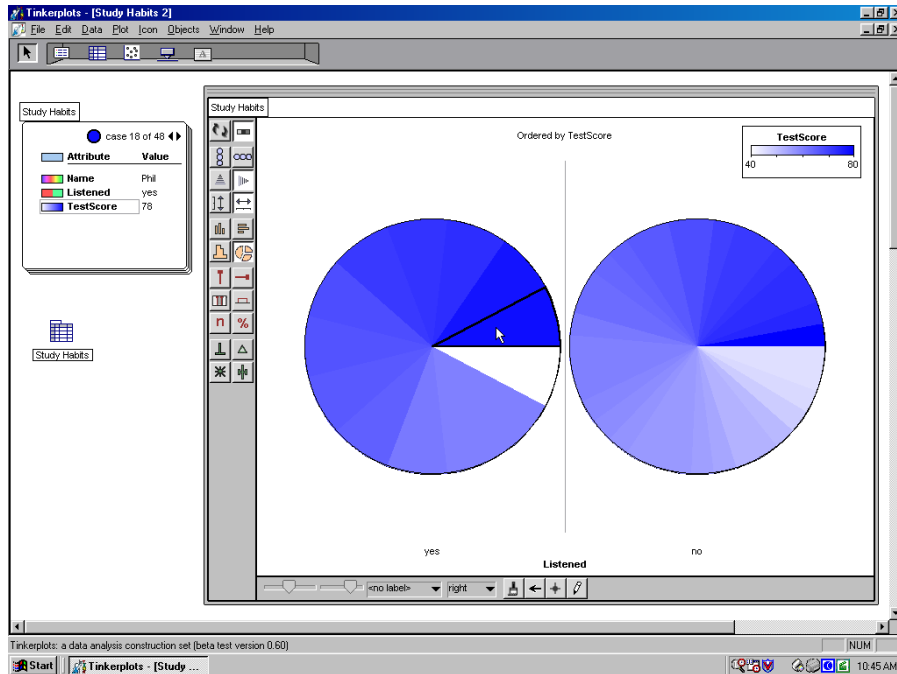


Figure 19. Side-by-Side Circle Graphs

Episode 5 illustrates how Brad and Joey analyzed the circle graphs they constructed.

### Episode 5

- 109 B: They look about the same to me.  
 110 J: No but see the triangles are like bigger here. (*The circle graph on the left.*)  
 111 B: So it still looks about the same.  
 112 J: This is more um, gradual. (*The circle graph on the right.*)  
 113 B: It doesn't matter. It doesn't matter.  
 114 J: Well look at the difference between that and that, and then that and that. (*Comparing adjacent sections in both graphs.*)  
 115 B: Because they have like 50 people.  
 116 J: They are quite similar. No...  
 117 B: They look about the same. They have the same amount of colors.  
 119 J: Woops. Shoot. (*Joey begins to modify the graph.*)  
 120 B: Yes, leave it like that Joey. Let's talk about this. (*The graph is the same as graph 5.*)  
 121 J: Ok um, did everyone get the same...  
 122 B: It looks about the same. I do not think it affected it. (*Listening to classical music did not affect test score.*)

The pair analyzed this graph based on color. However, Brad and Joey observed different characteristics of color. Brad insisted that the colors in both graphs are about the same (line 109, 111, 117). He thought that the change of color from white (light) to blue (dark) was about the same in both circle graphs. Joey recognized that the circle graph on

the right had a more gradual color fade from blue to white (line 110, 112, 114). Neither boy could convince the other of the validity of his point of view. Finally, Joey agreed that the graphs have about the same colors (line 121). The pair concluded that listening to classical music did not have an effect on test scores (line 122).

**Characterization of Pair 6C TinkerPlots session.** Table 6 is a chronological listing of how pair 6C’s session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

6C Critical/Creative Thinking	6C Statistical Thinking	Description
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 1	Produces side-by-side stacked dot plot
	Extreme/Outlier	Discusses highest score
Concretize		Discusses scores as grades
Causal Relationship		“but if you are naturally smart like me”
	Modifying Representation	Modifies side-by-side stacked dot plot
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 2	Produces side-by-side circle graphs
	Different Group Size	Notices difference in data set size
	Modifying Representation	Modifies side-by-side circle graphs
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 3	Produces side-by-side vertical dot plot
Causal Relationship		“This person is plain dumb.”
	Extreme/Outlier	Uses word “outlier”
Causal Relationship		“Sean lied about listening to classical music.”
Causal Relationship		“Laura she studied.”
Causal Relationship		“She probably listened to pop.” “She probably listened to rap.” “But this person Phil probably

		listened to popular.” “This guy probably listened to some Beethoven.”
Hypothesis		“It did not really effect.”
Verification		
Causal Relationship		“Maybe only the smart people listened to classical music.”
Re-statement of hypothesis		“would not effect it”
Modifying Representation		Modifies vertical dot plot
Representation 4		Produces red and green circle graph
Classify		Splits data into two columns
Sequence		Orders data from least to greatest
Causal Relationship		“This is a dumb person.”
Extreme/Outlier		Discusses highest score
Distribution Chunk		“there are more greens up here and more reds down here”
Modifying Representation		Modifies red and green circle graph
Classify		Splits data into two columns
Sequence		Orders data from least to greatest
Spread		“this is more um, gradual”
Spread		“they have the same amount of colors”
Conclusion		“I don’t think it effected it.”

*Table 6.* Chronological List of Codes for Pair 6C TinkerPlots Session

Brad and Joey used the following statistical thinking elements to help create and analyze their representations: classification, sequencing, modifying, extremes/outliers, distribution chunks, spread, and the recognition of different group sizes. The pair used creative and critical thinking elements to generate and verify their hypothesis, and add context to the data through the use of concretization and causal relationships.

One characteristic of Brad and Joey’s work during the session was the evolution of their thinking from the analysis of single data points to recognition of patterns within the entire data sets. Figures 15 and 17 were variations of dot plots with each student score represented by an individual circle. It seemed natural for the pair to click on the circles and determine a student’s test score. Brad and Joey focused solely on the highest and

lowest scores for both graphs. As opposed to this, Figures 18 and 19 were variations of circle graphs where all the student scores were arranged from least to greatest within their respective dataset. The circle graphs the pair created enabled them to shift their focus. With the circle graphs, they finally began to see the placement of scores in comparison to each other. The pair moved away from simply comparing individual test scores toward analyzing the entire distribution of data.

The boy's creation of graphs was not purposeful. They stumbled into the creation of new graphs by accident. Most of the time they would make modifications that drastically changed a graph and just continued the manipulations until they had created a new one. In the case where they switched from a dot plot to a circle graph Joey modified the graph in response to Brad questioning his hypothesis. He wanted to prove his hypothesis to Brad, but did not have a circle graph in mind when manipulating the dot plot.

The boys' work on this task can also be characterized by their use of the statistical thinking element spread. Makar and Confrey (2005) described spread as highlighting spatial aspects of the data. Spread of the data can be portrayed by nonstandard statistical language such as clustered, clumped, grouped, bunched, gathered, spread out, evenly distributed, scattered, and dispersed. The major component of analysis in Figure 19 was a discussion of the similarity in colors and their spread between the two graphs. The pair analyzed the color gradient, that is, the visible transitions from darker to lighter shades. Joey contended that the spread of the colors in the graph on the right was more "gradual" than the spread of the colors of the graph on the left. He interpreted this to mean that the scores for the students who listened to classical music were more evenly spread out,

whereas the scores for the students who didn't listen to classical music had one very low score, then primarily high scores. Brad believed the spread of the colors between the two graphs was about the same. He interpreted this to mean that listening to classical music did not affect test score. The characteristic of these graphs showing the entire data set as a whole helped them move beyond the analysis of single data points. These events signaled a turning point in the pair's analysis of the data.

Another characteristic of Brad and Joey's thinking throughout the session was the identification of causal relationships. Working through the task Brad and Joey became involved with the data to the point beyond just working with scores. They tried to explain the reasons for situations they encountered in the data. For example, in an effort to explain why a student received the score they did, the pair identified a relationship to the test score and the type of music listened to. At one point the pair had a playful exchange that related student's test scores to their own intelligence. For example, Brad stated that one of the students, Sean, who had the lowest overall test score "...is plain dumb." Joking around, Joey says "Like Brad." Brad responds "No like you." These exchanges personalized the raw data.

The boys' work on this task can be characterized by an analysis cycle that included analyzing the graphs they created, stating a hypothesis about which group of students did better on the Classical Music test, and attempting to verify the hypothesis. This cycle was repeated once during the session. What prompted this cycle was when Brad questioned Joey's hypothesis that listening to classical music did not affect test scores, asking "Are you sure?" and "Why is that so?" which in turn, provoked Joey to try to justify his hypothesis.

The significance of this is that in their analysis of Figure 19 both of the boys' changed their hypothesis. Brad insisted that the change of color from white to blue in both graphs was about the same, while Joey recognized that one of the circle graphs had a more gradual color fade. Neither of them could convince the other of the validity of his point of view. Finally, Joey ended up conceding to Brad's view that listening to classical music did not have an effect on test scores even though he was not in complete agreement.

In summary, Brad and Joey's work in the TinkerPlots was characterized by their use of spread, their identification of causal relationships, their evolution in thinking based on the type of graphs they were analyzing, and their analysis cycle.

**Differences in thinking when working in the two mediums.** This section describes the differences in the pair's thinking between the two mediums. A notable difference between the TinkerPlots task and the paper and pencil task was that during the TinkerPlots task Brad and Joey created a number of different graphs and used a variety of strategies to analyze the data. This was never observed to be the case during the paper and pencil task since they only inspected the data visually and did not take the entire data set into account, only writing down the extreme scores.

At the beginning of both the paper and pencil task and the TinkerPlots task, the strategies Brad and Joey used were similar. At the beginning of the paper and pencil task, the pair's discussion and analysis focused on identifying then comparing the highest and lowest test scores for both data sets. Although Brad and Joey began the TinkerPlots session using a similar approach, their thinking began to evolve as they worked their way through the task. Their last two graphs were variations of circle graphs where all the student scores were arranged from least to greatest within their respective dataset. The



circle graphs the pair created enabled them to shift their focus; they began to analyze the entire data set, discussing the transition of colors between the data sets and comparing the placement of scores.

The pairs work in both mediums can be characterized by an analysis cycle that included stating a hypothesis about which group of students did better, and attempting to verify the hypothesis. In both environments this cycle was repeated once and the pair's hypothesis was that Driver's Education or listening to Classical Music did not affect test scores. In paper and pencil their analysis cycle can be characterized as limited, as they only discussed the extremes and re-stated their hypothesis. In TinkerPlots the pair's analysis cycle was more complex. The creation of additional graphs helped them further analyze the data and verify their hypothesis. During the analysis cycle Joey tried to convince Brad that listening to classical music had an effect on test scores. Their disagreement over this is what prompted them to continue their analysis. Much of their contention stemmed from a difference of opinion over the interpretation of Figure 19, side-by-side circle graphs. The pair analyzed the color gradient, that is, the visible transitions from darker to lighter shades. Joey contended that the spread of the colors in the graph on the right was more "gradual" than the spread of the colors of the graph on the left. Brad believed the spread of the colors between the two graphs was about the same. The characteristic of this graph showed the entire data set as a whole, helping them move beyond the analysis of single data points.

At the end of the TinkerPlots session Joey conceded to Brad's view that listening to classical music did not have an effect on test scores even though he was not in complete agreement. However, in paper and pencil both partners' were both in agreement

regarding the hypothesis.

In both environments, Brad and Joey personalized the data by concretizing and identifying causal relationships. A major difference in their personalization of the data was the number of times during the two sessions that they demonstrated these thinking processes. In paper and pencil the pair demonstrated one instance of concretizing and identified one causal relationship. In TinkerPlots personalization of the data took on a more significant role. Brad and Joey demonstrated the use of causal relationships seven times, with these being notably longer back and forth exchanges. Their identification of causal relationship was purposeful as it aided them in becoming involved with the data beyond just working with scores.

In paper and pencil the pair concluded that Driver's Education did not have an effect on test scores. The identification of causal relationships afforded the pair to opportunity to move beyond this conclusion, as they tried to explain the reasons for situations they encountered in the data. For example, the pair explained the effect of listening to a certain type of music might have on a student's score. Joey said, Laura "she probably listened to pop," and Brad says "She probably listened to rap or something good" (episode 3, lines 60 – 61). As for Sean, Brad stated that he got an F because he "probably listened to some Beethoven" (line 63).

There were differences in how the pair dealt with the different group sizes in both environments. In paper and pencil, the pair did not recognize this difference – it was not visible in their representation of the data. In TinkerPlots, the pair recognized that the two groups were comprised of a different number of students. This observation played a part in their analysis of the graphs they created. It aided them in their analysis of the spread of

the data in their circle graphs and also in being able to compare the placement of distribution chunks to each other between the data sets.

### **Pair 7A**

**Summary of Pair 7A paper and pencil session.** Pam and Susan began the Driver's Education task (see Appendix C) by separating the cards into two respective stacks: those who took Drivers' Education and those who did not. The pair decided to compare the scores of the students who took Drivers' Education to the students who did not by pairing a card from one group with a card from the other group. They randomly drew one card from one stack and one card from the other stack and compared the scores. Then, on a sheet of paper, they created a two-column chart to record the scores: one column was reserved for those that took Drivers' Education and the other was for those that did not. On this chart they recorded the name and score of the higher score of the pair in its respective column. After recording the scores for three pairs of students they decided this method would not work because they noticed there were more score cards in the pile of students who did not take Driver's Education. They abandoned this strategy and Pam suggested another approach.

Pam suggested that they should create two fractions: one for the students who took Drivers' Education and one for the students who did not. The numerator of each fraction would be the total number of points that each group scored whereas the denominator would be the total points possible for that group. These fractions would represent the number of points that a group of students scored out of the total number of possible points. Once the fractions were created they would then be converted to percentages. Pam reasoned that they could then simply compare the respective

percentages to determine which group of students performed better on the test.

Susan agreed with this approach and the pair began to create the two fractions. While Susan performed the calculations for the people that took Drivers' Education, Pam did the same for the people who did not take the course. They began by calculating the denominator of each fraction: they found the total number of points possible for each group of students by multiplying 70 (total points possible on the test) times the number of students in each group. Next, they calculated the numerator for each fraction summing the scores for their respective group of students. The two fractions were  $841/910$  for the students who took Drivers' Education and  $1798/2150$  for the students who did not. However, the pair made mistakes in their calculation of the two fractions. The denominator of the fraction for the students who did not take Drivers' Education was incorrect: 70 times 35 is 2450 not 2150. Also, the numerators of both fractions were incorrect. The numerator for the students who took Drivers' Education should have been 728 and the numerator for the students who did not take Drivers' Education should have been 1788.

After Pam and Susan created their two fractions, and according to the strategy they discussed and agreed to earlier, they would have converted the two fractions to percentages. At this point, however, the pair forgot about their initial strategy and began to discuss again how they should compare the two fractions (episode 1, line 1). Episode 1 illustrates how Pam and Susan analyzed the fractions they constructed.

#### Episode 1

- 54 S: We could make it like a fraction, and then like narrow it down.  
55 P: We could do that yeah, and then like try to find the common denominator.  
56 S: Yeah.  
57 P: Well you could, no, you divide by 10, yeah.

- 58 S: Would it be a even number?  
 59 P: It would be, ur 91.  
 60 S: It wouldn't be a number then?  
 61 P: Wait, what?  
 62 S: You can divide. What would that be divided by 10? (*Susan points to 841.*)  
 63 P: It's me, maybe 84.1... (*Divides 841 and 910 by 10 mentally.*)  
 64 S: Uh, huh.  
 65 P: I think... it would be 84.1. So then you have 84.1 over 91. Then you would have...  
 66 S: Then what about dividing that?  
 67 P: 1, 2, and then 1, 4, 8. (*Checks division with multi-digit multiplication:  $841 \times 100 = 841.00$ .*) No, yea it would be 84.1. So then you have like 84.1 over 91. Then you would have...  
 68 S: 84 divided by...  
 69 P: 7, 1, 179.8 over 215. (*Divides 1798 and 2150 by 10 mentally.*)  
 70 S: Uh, huh.  
 71 P: Is that right? Yeah.

Susan suggested they write the fractions in simpler terms (line 54) and Pam agreed, suggesting that they find a common denominator (line 55). Susan worked with the students who took Drivers' Education while Pam's efforts were directed toward the students who did not. They each divided their fraction by 10 (lines 57 – 69). The pair completed multiple iterations of the simplification process dividing their fractions by numbers they were easily divisible by such as 2, 5 and 10. Much of their work was mental calculation, combined with the multiplication and long-division algorithms. Many of their cycles of simplification were accurate and completed with ease.

During other cycles the pair made many errors in their mathematical calculations and applied questionable mathematical techniques in their simplification process. However, some of the calculations were performed using mental reasoning without the use of algorithms, demonstrating Pam and Susan's flexibility of thought.

About half-way through the process of simplifying the fractions, Susan recognized that they needed to work toward obtaining a common denominator. At one

point Susan lost hope that they would ever find a common denominator for the two fractions and pleaded – “let’s make sure it is getting closer.” To assure Susan that the fractions were “getting closer” to one another, Pam subtracted the numerator and denominator of the last two fractions they created (21.50/22.75 and 10.5/11.375). Since 22.75-21.50 yielded a larger number (1.25) than 11.375-10.5 (.875), Pam assured Susan that the fractions are getting a lot closer to one another

After a few more cycles of simplification, the final fractions the pair produced were 5.25/5.68 for the students who took Drivers’ Education and 4.59/5.375 for the students who did not. Since the denominators of the two fractions were within one point of one another Pam and Susan reasoned that they were close enough in value to compare. They stopped their simplification process and began their analysis. Episode 5 demonstrates Pam and Susan’s analysis of the fractions they constructed.

#### Episode 5

- 263 P: What if you just took off your decimals? And I took off my decimals. (*Pam crosses-out the decimals on the fraction 4.59/5.375 and 5.25/5.68.*) Yours is 100%, mine is 80%. (*One fraction is now 5/5 or 100% and the other fraction is 4/5 or 80%.*)
- 264 S: Mine cannot be 100% though because a lot of the people didn’t get.
- 265 P: But yours if you look.
- 266 S: When the denominator gets down to this small from where they were, even the little decimal numbers count.
- 267 P: That’s true, because that would be like .4 and mine would be .8. (*They were finding the difference between the numerator and denominators: 5.375-4.59 is approximately .8 and 5.68-5.25 is approximately .4.*)

To analyze the fractions the pair then eliminated the decimal portion of the fractions and compared 4/5 to 5/5 (line 263). Their interpretation was that 4/5 represented 80% and 5/5 represented 100% (line 263). Susan saw that this could not be true; many of the people who took Drivers’ Education did not score 100% on the test (line 264). She

explained that “When the denominator gets down to this small from where they were, even little decimal numbers count” (line 266), so eliminating the decimal portions of the fractions will not work.

At this point, the pair was out of allotted time for the session. The researcher asked them what their conclusion was. They responded that the students who took Drivers’ Education did better on the test. For the fraction  $4.59/5.375$ , they reasoned that the numerator of 4.59 was approximately 8 tenths of a point less than the denominator of 5.375. For the fraction  $5.25/5.68$ , the numerator 5.25 was approximately 4 tenths of a point less than the denominator 5.68. The numerator of the first fraction was closer in value to the denominator; therefore the fraction  $5.25/5.68$  was larger than the fraction  $4.59/5.375$ .

**Characterization of Pair 7A paper and pencil session.** Table 7 is a chronological listing of how pair 6A’s session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

7A Critical/Creative Thinking	7A Statistical Thinking	Description
	Classifying	Sorts data cards into two piles
	Different Group Size	“There are more people in this pile.”
	Modifying Strategy	Changes representation strategy
	Average	Uses word “average”
	Representation	Constructs fractions $841/910$ and $1798/2150$
	Modifying Representation	Simplifies fractions to $5.25/5.68$ and $4.59/5.375$
	Modifying Representation	Converts fractions to 80% and 100%
Conclusion		The students who took Drivers’ Education did better on the test.

*Table 7. Chronological List of Codes for Pair 7A Paper and Pencil Session*

Pam and Susan exhibited the following aspects of statistical thinking as they worked on the Driver's Education task: representation, classifying, average, modifying, and recognition of different group size. The only critical or creative thinking element used during the session was hypothesizing. The pair only produced one hypothesis, and it was the conclusion.

Pam and Susan's work on this task can be characterized by the fractions they created that represented the average score for both group of students. The pair created two fractions: one for the students who took Drivers' Education and one for the students who did not. The numerator of each fraction was the total number of points that each group scored whereas the denominator was the total points possible for that group. These fractions were converted to percentages and then compared for size.

The pair appeared to recognize that the percentages could be compared even through the group sizes were different. The first method the pair used to compare the two groups of students was not successful due to this. The pair's second strategy, which lumped the data for both data sets into an aggregate score made it possible to compare the two groups of students on an equal basis; there were no longer unequal group sizes to deal with.

Another characteristic of Pam and Susan's work was how they turned the task into a computational exercise of comparing two numbers. Pam and Susan originally planned on converting their fractions into percentages. They forgot about this strategy and decided to find a common denominator for the two fractions. They did not have a plan on how to find a common denominator. They haphazardly began dividing the fractions with numbers that were easily divisible, reducing the fractions to lowest terms.



They believed that reducing the fractions to simplest terms would yield a common denominator. This was a very long and drawn out process that left little time for analysis. So long, that the researcher had to step in and ask for their conclusion.

In addition, the girls' made many mistakes in their calculation of the two fractions. The errors that the pair made could have sent them in the wrong direction and they could have based their decision on incorrect mathematical procedures. Fortunately, their mistakes "evened out" and the fractions were roughly proportionate. Their final version of the fractions was  $4.59/5.375$  for the students who did not take driver's education and  $5.25/5.68$  for the students who took drivers education. They eliminated the decimal portion off of fractions, leaving  $4/5$  and  $5/5$ . Their interpretation of the size of the two fractions was that  $4/5$  represented 80% and  $5/5$  represented 100%.

Pam and Susan arrived at one response in the paper and pencil task. They only gave one hypothesis, and it was the conclusion. Subsequently, they made no effort to continue their analysis or verify their response after reaching their conclusion. They did not discuss the questions they were asked to answer during any part of the session, until the very last moment.

In summary, Pam and Susan's work in the paper and pencil environment was characterized by the fractions they created, the computational exercise it took to create the fractions, their comparison of the fractions using percents, and the difference in group size.

**Summary of Pair 7A TinkerPlots session.** The second task Pam and Susan completed required them to determine whether students studying for a mathematics test while listening to classical music performed better than those who did not (see Appendix

B). The data in TinkerPlots were presented to the participants in a random arrangement of 48 circles in differing shades of blue. Each circle represented a student and his or her score on the mathematics test. Each circle was filled with a shade of blue that reflected the student's test score. Darker shades of blue represented higher scores whereas lighter shades of blue represented lower scores. Refer to Appendix A for an image of the way the data were presented to the students.

Pam and Susan began the session by creating a variation of a dot plot that was separated into two columns along the horizontal axis and rows on the vertical axis (Figure 20). The left-hand column displayed the test scores of the students who had listened to classical music while studying and the right-hand column displayed the test scores of those students that did not listen to classical music while they studied. The vertical axis was segmented into sections of test scores (bins) in increments of five points each. Within each bin individual student test scores, represented by rectangles (bars), were arranged in horizontal rows. The height of each rectangle represented the student's test score. The taller the rectangle was the higher the student score. For example, the rectangles in the topmost bin were taller than the rectangle in the lowest bin. There was even variation in the height of rectangles in each bin, the taller the rectangle within a bin, the higher the student score.

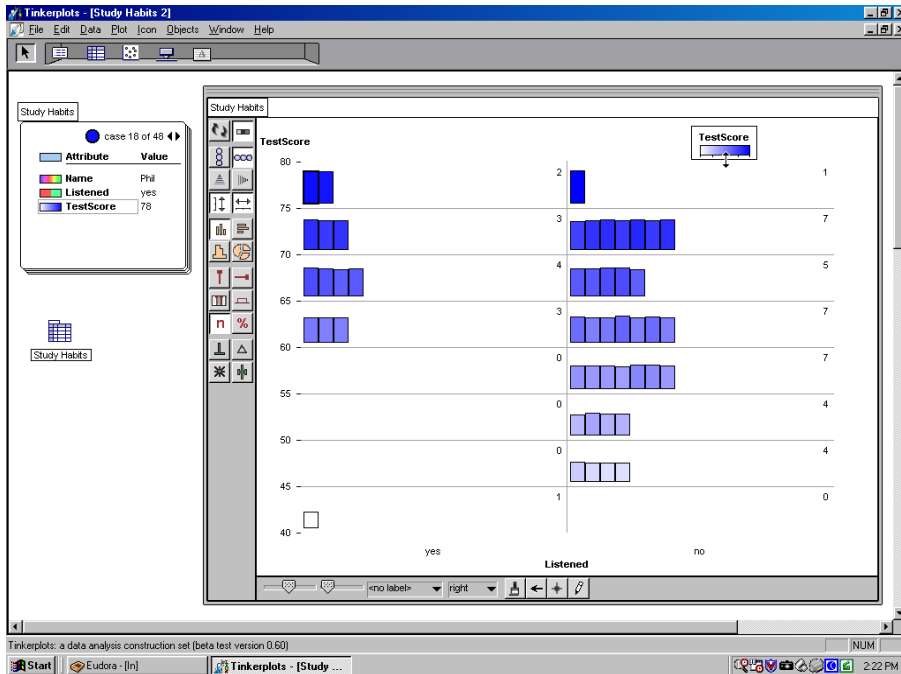


Figure 20. Side-by-Side Vertical Dot Plot with Bins

Episode 1 illustrates how Pam and Susan analyzed the side-by-side vertical dot plot with bins they had created.

Episode 1

- 25 S: I guess the people that are up higher, said no, did worse. Or usually do that good, 'cos didn't they didn't change anything. And that those people all got high scores.
- 26 P: Right, so, what if you, do we need that you think.
- 27 S: No
- 28 P: Um, ok, so how did the scores, test scores of students that listened to classical music compare to the test scores of students that did listen to classical music while studying?
- 29 S: Um...
- 30 P: Then let's see, I'd say that it looks like the kids who listened to classical music overall did better on the test.
- 31 S: Yeah.
- 32 P: But is there like as many people? There is not.
- 33 S: There is not as many people but...
- 34 P: But still the people who...

While working on the task Pam and Susan suggested causes of an event or situation. In line 25 Susan stated that “I guess the people that are up higher, said no, did worse. Or

usually do that good, ‘cos didn’t they didn’t change anything. And that those people all got high scores.” What Susan was referring to were the students who did not listen to classical music but still scored high on the test. These students scored high, but did not change any of their normal study habits. These students usually do well on exams; this one was no exception. This was her way of explaining why students who did not listen to classical music also scored high on the test.

Pam hypothesized that the students that listened to classical music while they studied scored higher overall, recognizing that the students who listened to classical music were in the higher bins on their side of the graph (line 30). The pair recognized that there were a different number of student in both groups (line 32). This observation led the pair to question their hypothesis and they continued analysis to verify their claim.

To confirm their response, Susan suggested that they should find the average of the test scores for both groups of students. From the TinkerPlots drop down menu, they chose the option “show value of average.” This option was disabled because another option, “Show location of mean,” was not selected previously. I told them depending on the graph and the options that were selected, other options may not be available. Susan changed the bars to circles thinking this would allow them to show the average. They checked the drop down menu again and “show value of average” was still not available. Pam said that she thought it had to be a bar graph for the option to work.

In response to this the pair created a bar graph (Figure 21) similar to Figure 20; the only difference was that the axes were swapped. In both graphs, the rectangular bars had the same orientation. The height of each rectangle represented the students test score.

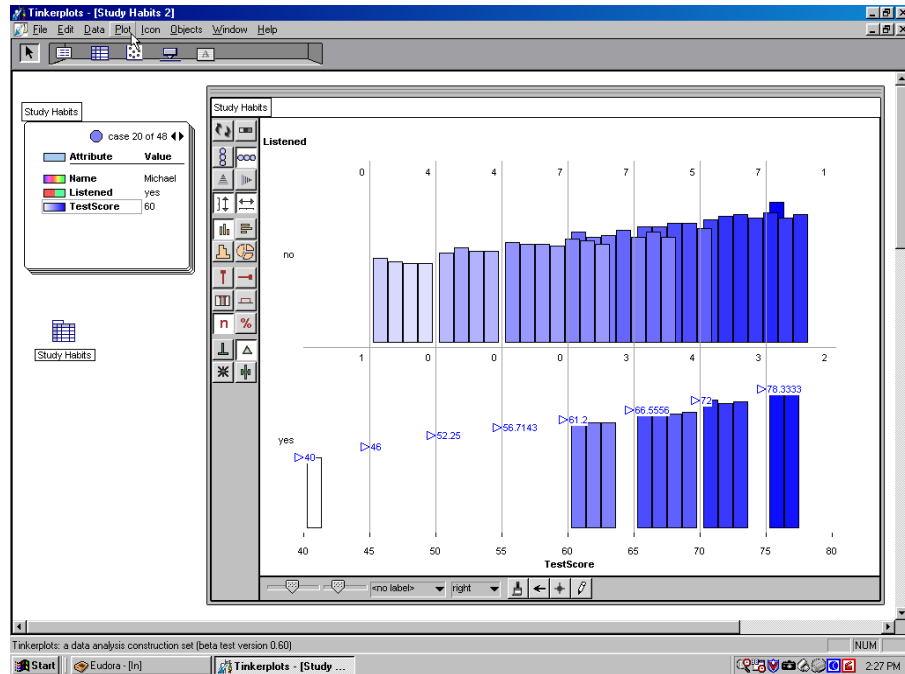


Figure 21. Side-by-Side Horizontal Bar Chart with Bins

On the graph shown in Figure 21, the pair was able to display the average. However, the average that was displayed represented the average for each column of the entire graph and thus included both groups of students. The pair recognized this average would not help them compare the test scores of the two groups of students. Pam suggested that they reorganize the graph by swapping axes. The result was another graph (Figure 22) similar to Figure 20, except that the student test scores were represented by circles instead of rectangular bars. During this process, the pair forgot about finding the average and began to analyze the graph without attempting to find out what the average was.

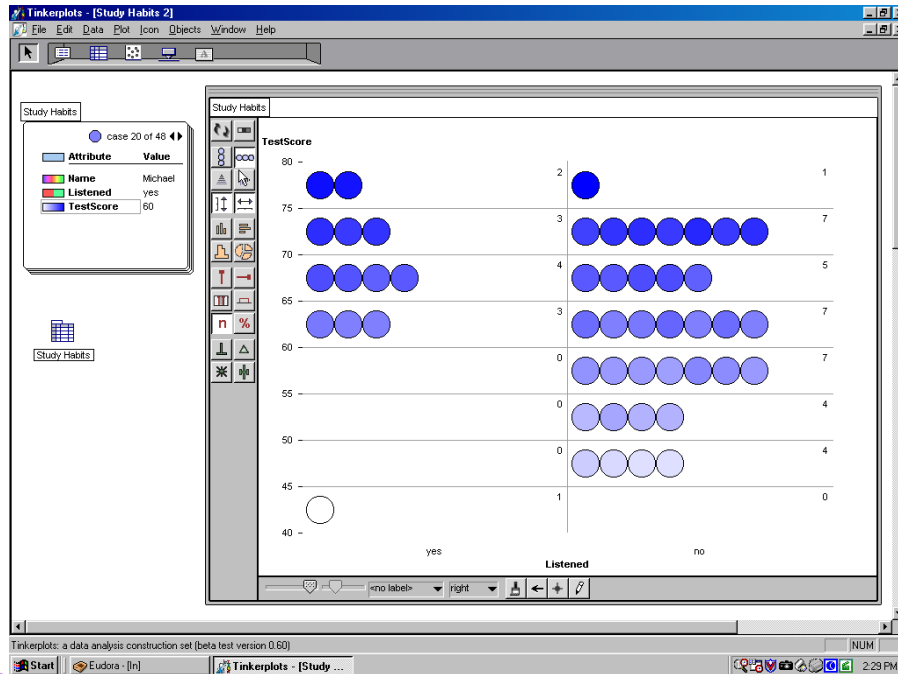


Figure 22. Side-by-Side Vertical Dot Plot with Bins

Episode 2 illustrates how Pam and Susan analyzed the side-by-side vertical dot plot they constructed.

Episode 2

- 107 P: Then, so the kids who listened to classical music overall did a lot better on the test. And so in conclusion you could say that if you listened to classical music you will do better on the classical music test.
- 108 S: Not necessarily though. It just kind of depends on the kid. Because there was more people that didn't listen to it that... 75. But well yeah...
- 109 P: I don't know...
- 110 S: I guess if you did listen, if you listened to music...
- 111 P: But look at all the kids.
- 112 S: Well I guess you could say that if you listened to music you're gonna do good, but if you don't listen to music you are not necessarily going to do bad.
- 113 P: You could say that if you listened to classical music you will do...
- 114 S: Good
- 115 P: You'll do good on the test, and if you listen to classical music...
- 116 S: If you don't...
- 117 P: If you don't listen to classical music...
- 118 S: You are not going to do bad necessarily.
- 119 P: You are not going to do as well.
- 120 S: You are just do your average.
- 121 P: You are just going to what is normal for you.

122 S: yea. Ok.

Pam hypothesized that the students who listened to classical music scored better overall on the test (line 107). Susan provided more detail about how she viewed the data. In response to Pam’s general response, she added that “Not necessarily though. It just kind of depends on the kid” (line 108). Pam refined her hypothesis stating “Well I guess you could say that if you listened to music you’re gonna do good, but if you don’t listen to music you are not necessarily going to do bad” (line 112). Finally, Susan concluded that you will “just do your average” (line 120), and Pam conceded “You are just going to do what is normal for you” (line 121).

**Characterization of Pair 7A TinkerPlots session.** Table 8 is a chronological listing of how pair 7A’s session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

7A Critical/Creative Thinking	7A Statistical Thinking	Description
	Outlier/Extreme	Refers to lowest score
	Classifying	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 1	Produces side-by-side vertical dot plot
Causal Relationship		“They didn’t change anything.”
	Distribution Chunk	Discusses scores that are up higher on graph
Hypothesis		“Kids who listened to classical music, overall did better on the test.”
	Different Group Sizes	Notices difference in data set size
Verify	Average	Tries to display average test scores on graph
	Modifying Representation	Modifies side-by-side vertical dot plot
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 2	Produces side-by-side horizontal bar chart
	Average	Displays average of test scores for

		each column on graph
	Modifying Representation	Modifies side-by-side horizontal bar chart
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 3	Produces side-by-side vertical dot plot
Re-statement of hypothesis		“The kids who listened to classical music did a lot better on the test.”
Addition to hypothesis		“It kind of depends on the kid.”
Re-statement of hypothesis		“If you listen to classical music you’re gonna do good, but if you don’t listen to classical music you are not necessarily going to do bad.”
Conclusion		“You will just do your average.”
Conclusion		“You are just going to do what is normal for you.”

*Table 8.* Chronological List of Codes for Pair 7A TinkerPlots Session

The statistical techniques that Pam and Susan used during the session were classification, sequencing, modifying, extremes/outliers, average, distribution chunks, and the recognition of different group sizes. The pair used creative and critical thinking elements to generate and verify their hypothesis, and add context to the data through the use a causal relationship.

The girls’ work on this task can be characterized by an analysis cycle that included analyzing the graphs they created, stating a hypothesis about which group of students did better on the Classical Music test, and attempting to verify the hypothesis. This cycle was repeated once during the session. The pair’s first hypothesis was generated during their analysis of Figure 20. Toward the end of this analysis they recognized the difference in group sizes which prompted them to create new graphs so they could further analyze the data. In their analysis of Figure 21 the pair was unable to determine the average test score for both groups of students. In an effort to determine the average test scores the pair created Figure 22. However, during this process the pair



forgot about finding the average and began to analyze the graph without attempting to find out what the average was. They relied on visual inspection for their analysis of Figure 22.

An outcome of the hypothesis cycle was the refinement of their hypothesis. Pam and Susan's version of the conclusion were slightly different. Pam hypothesized that "The kids who listened to classical music overall did a lot better on the test." During the analysis of their three graphs, Pam identified distribution chunks that helped her observe that the students who listened to classical music were higher on their side of the graph than the students who did not listen to classical music.

Susan provided more detail about how she viewed the data. In response to Pam's hypothesis, she added that "Not necessarily though. It just kind of depends on the kid." Susan was able to see the entire range of the data arranged from least to greatest, and how the scores of both groups of students were placed throughout. She looked at the entire distribution of students who did not listen to classical music and recognized that some of these students scored high also. She included these students in her hypothesis, adding that "if you do not listen to classical music you are not necessarily going to do bad" you are going to do what is "average" or "normal" for you. Figure 22 provided a visual where Susan could see the entire range of scores for students who did not listen to classical music sequenced from least to greatest. This visual encouraged her to include this group of students in her conclusion.

A significant part of the pair's analysis was that they were unable to find the average of the test scores in Figure 21. In an effort to determine the average test scores the pair created Figure 22. However, during this process the pair forgot about finding the

average and began to analyze the graph without attempting to find out what the average was. They relied on visual inspection for their analysis of Figure 22.

In summary, Pam and Susan's work in TinkerPlots was characterized by their repeating the analysis one time, their inability to find the average, the refinement of their hypothesis, their visual analysis of the graphs, and observation of distribution chunks.

**Differences in thinking when working in the two mediums.** This section describes the differences in the pair's thinking between the two mediums. In paper and pencil the pair represented the data with two fractions. In TinkerPlots they created three different graphs.

In paper and pencil, Pam and Susan did not demonstrate the elements of hypothesizing and verifying involved in the analysis cycle. They only gave one hypothesis, and it was the conclusion. They made no effort to continue their analysis or verify their response after reaching a conclusion. In TinkerPlots the analysis cycle was repeated once during the session. Pam and Susan came to a hypothesis early on, after a brief analysis of the first of three graphs that the pair constructed. Pam and Susan recognized that there were disproportionate group sizes which caused them to question their response. This triggered the pair to attempt to verify their initial hypothesis, and in response they created a different graph. In their analysis of this new graph the pair was unable to determine the average test score for both groups of students. In an effort to do so they created another new graph. However, during this process the pair forgot about finding the average and began to analyze the graph without attempting to find out what the average was, relying on visual inspection for their analysis.

An outcome of the pair's hypothesis cycle in TinkerPlots was the refinement of their hypothesis. The pair's conclusion was that "The kids who listened to classical music overall did a lot better on the test." Susan provided more detail about how she viewed the data. In response to Pam's hypothesis, she added that "Not necessarily though. It just kind of depends on the kid."

In TinkerPlots, Susan was able to see the entire range of the data arranged from least to greatest, and how the scores of both groups of students were placed throughout. She looked at the entire distribution of students who did not listen to classical music and recognized that some of these students scored high also. She included these students in her hypothesis, adding that "if you do not listen to classical music you are not necessarily going to do bad" you are going to do what is "average" or "normal" for you. Whereas, Pam identified distribution chunks that aided her observation that students who listened to classical music scored higher in general.

In paper and pencil the pair worked the entire session on simplifying the two fractions they had created to lowest terms, believing that this process would yield a common denominator. They turned this task into a computational exercise of comparing two numbers that was a very long and drawn out process that involved multiple iterations of simplification. There were many mistakes in their calculation of the two fractions. Although the pair arrived at a reasonable conclusion, part of their logical reasoning was weakened by faulty mathematical procedures.

A significant part of the pair's analysis in TinkerPlots was that they were unable to find the average of the test scores in the side-by-side horizontal bar chart with bins.

Despite not being able to calculate the measures of center, the pair was able to overcome this by using the graphing capabilities of TinkerPlots to create new graphs quickly and efficiently. Due to this, they were able to continue their analysis and reach a reasonable conclusion.

There were differences in how the pair dealt with the different group sizes in both environments. In paper and pencil the first method the pair used to analyze the data was unsuccessful due to their inability to compare groups with different sizes. For their second strategy, they appeared to recognize that the percentages could be compared even though the group sizes were different. In TinkerPlots, the pair understood there was a difference in group sizes and decided to verify their original hypothesis and continue analysis based on this. The visual of different group sizes was a prominent feature in TinkerPlots. It aided Susan in adding complexity to her conclusion with a response that was richer in context.

### **Pair 7B**

**Summary of Pair 7B paper and pencil session.** Camille and Larissa began the Driver's Education task (see Appendix C) by tallying up the number of people that took Drivers' Education and the number of people that did not take Drivers' Education. They found that 15 people took drivers education and 35 people did not take drivers education. The pair had a discussion over what constituted a "good" grade. The pair decided that a score of 50 (70 was the total possible points on the test) would be a good score since it is close to a B. They put a star beside the students who got a good score on the data sheet. This helped them count the number of students with scores above and below 50.

Camille and Larissa used these counts to create fractions out of the total number of students (the total number of students was 50). On a piece of paper they created two columns – one labeled “Did” for did the students that took drivers education, and the other labeled “Didn’t” for the students who did not take drivers education (Figure 23). Under the heading “Did” they wrote 13/50 and under the heading “Didn’t” they wrote 35/50.

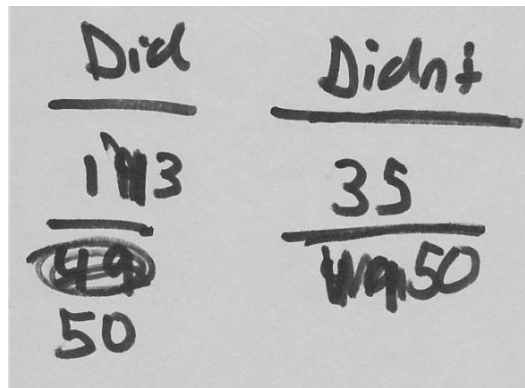


Figure 23. Fractions for Students Scoring Over 50

Without any analysis, the pair abandoned their work represented in Figure 23 and continued to represent and analyze the data. Earlier in the session they had determined that 50 was a good score. Based on this criteria, they counted the number of students in each group and used this as the denominator for students who scored above and below 50. On the same sheet of paper that the original fractions had been recorded, in a different area, Larissa wrote four headings: Good Course, Bad Course, Good no course, and Bad no course (Figure 24). These four categories represented students who took the course and did good (Good Course), took the course and did bad (Bad Course), did not take course and did good (Good no course, and did not take course and did bad (Bad no course). Figure 24 shows the headings and fractions they created for each of the

categories.

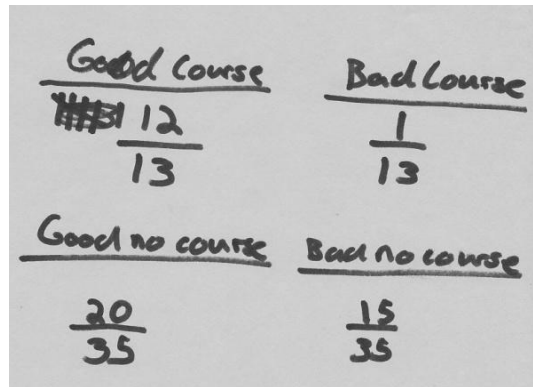


Figure 24. Fractions for Students Scoring Above and Below 50

Camille also created a representation with the same data (Figure 25). On a different sheet of paper, Camille created two columns: "Did" for the students who did take the course and "Didn't" for the students who did not take the course. Under "Did," Camille wrote "Good= 12 of 13" and "Bad= 1 of 13." Under the column "Didn't" she wrote "Good=20 of 35" and "Bad=15 of 35" (Figure 25).

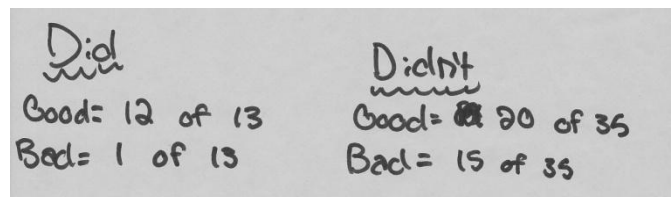


Figure 25. Fractions for Students Scoring Above and Below 50

Without analysis of their representations, the pair generated a hypothesis (Episode 1, Lines 81 - 82). Episode 1 is the discussion that led to their conclusion.

#### Episode 1

- 81 C: Okay I think that taking the course helped.  
82 L: I think it helps but I still think you can still pass the class if you do not take it.  
83 C: Cos' me too. So...  
84 L: So what is...  
85 C: It helps.  
86 L: So for 1... okay... (Referring to question 1.)  
87 C: So like...

- 88 L: Okay that was for 1 this is for 2... What conclusion... okay so I think yeah it helps to take the test. (*Referring to question 2.*)
- 89 C: Course
- 90 L: Course
- 91 C: But you don't need it to pass...
- 92 L: Uh huh
- 93 C: And you can still pass without it.
- 94 L: Okay

Camille hypothesized that taking the course helped (line 81). Larissa agreed but added a caveat to Camille's hypothesis saying that "I think it helps but I still think you can still pass the class if you do not take it (line 82). Camille re-stated her conclusion making it the same as Larissa's (lines 91, 93).

**Characterization of Pair 7B paper and pencil session.** Table 9 is a chronological listing of how pair 7B's session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

7B Critical/Creative Thinking	7B Statistical Thinking	Description
Concretize		Discusses good grades
Concretize		Discusses good grades
	Outlier/Extreme	Discuss high score
Concretize		Discuss good and bad grades
	Classify	Counts students in both data sets
	Different Group Size	Notices difference in data set size
	Representation 1	Produces two fractions for scores over 50
Concretize		Discusses good and bad as grades
	Representation 2	Produces four fractions for scores above and below 50
Hypothesis		"Taking the course helped."
Conclusion / Addition to Hypothesis		"I think it helps but I still think you can still pass the class if you do not take it."

*Table 9.* Chronological List of Codes for Pair 7B Paper and Pencil Session

In this section, I characterize the work of Camille and Larissa during the paper

and pencil session, specifically with regard to their use of critical, creative, and statistical thinking. The pair created four fractions to represent the data. In the creation of the fractions they used the critical and creative thinking elements concretization and hypothesizing. In addition, the statistical thinking elements used were outlier/extreme, classification, and recognition of different group sizes.

Camille and Larissa's work on this task can be characterized by the technique they used to create and compare the fractions they constructed. They employed a method which I called the "cut-off" technique. The pair decided to use a score of 50 to be the cut-off between "good" and "bad" scores. They decided that 50 points would constitute a good score since it was close to a B. No scientific method was used to determine this.

The number of students who scored above and below the cut-off were counted and included in the numerator of the fraction for their respective group. Then these counts were placed over the total number of students in the group. Both Camille and Larissa recorded the fractions on two different sheets of paper. The representations were created with the same data, but each team member created their own rendition of it.

The girls' work on this task can also be characterized by the lack of an analysis cycle. The pair spent most of the session constructing the fractions and very little time analyzing them. Camille and Larissa both presented a hypothesis, with Larissa's hypothesis being accepted as their final conclusion. They made no effort to continue their analysis or verify their response after reaching their conclusion.

The pair did little to analyze the representations they created. The two fractions representing the students who did "good" in both groups were compared. The major piece of their analysis was based on Larissa's statement that "12 of the 13 people that took the



test did good and 20 of the 35 people that did not take the test passed – did good.” This contributed to the difference in Camille and Larissa’s hypotheses. Camille hypothesized that taking Drivers’ Education helped. Larissa provided more detail about how she viewed the data. In response to Camille’s hypothesis, she added that “I think it helps but I still think you can still pass the class if you do not take it.” Larissa focused on the students who did “good” in each group without comparing the size of the fractions to one another. She only noticed that there were students from both fractions that scored over 50 points on the test.

Another characteristic of the Camille and Larissa’s work in the session was their use of concretization. The pair made many references relating student test scores to letter grades. Their discourse throughout the session consisted primarily about whether or not students did good or not; specifically whether the students scored over 50 points, which was considered the cut-off for a good and bad score.

In summary, Camille and Larissa’s work in the paper and pencil environment was characterized by their use of the “cut-off” technique to create their fractions, the lack of an analysis cycle, and their use of concretization.

**Summary of Pair 7B TinkerPlots session.** The second task Camille and Larissa completed required them to determine whether students studying for a mathematics test while listening to classical music performed better than those who did not (see Appendix B). The data in TinkerPlots were presented to the participants in a random arrangement of 48 circles in differing shades of blue. Each circle represented a student and his or her score on the mathematics test. Each circle was filled with a shade of blue that reflected the student’s test score. Darker shades of blue represented higher scores whereas lighter

shades of blue represented lower scores. Refer to Appendix A for an image of the way the data were presented to the students.

Camille and Larissa began the TinkerPlots session by sorting the test scores into two bins. The bin that represented the students who listened to classical music while they studied was labeled “yes” and the bin for the students that did not listen to classical music was labeled “no”. The pair quickly modified this graph creating several different versions until they arrived at the side-by-side vertical dot plot, illustrated in Figure 26, which they analyzed.

Figure 26 was separated into two columns along the horizontal axis with test scores on the vertical axis. The left column displayed students that had listened to classical music and the right column displayed those students that had not listened to classical. The vertical axis was segmented into sections of test scores in increments (bins) of eight points each. Within each bin, student test scores, represented by circles, were arranged in horizontal rows, ordered by test score.

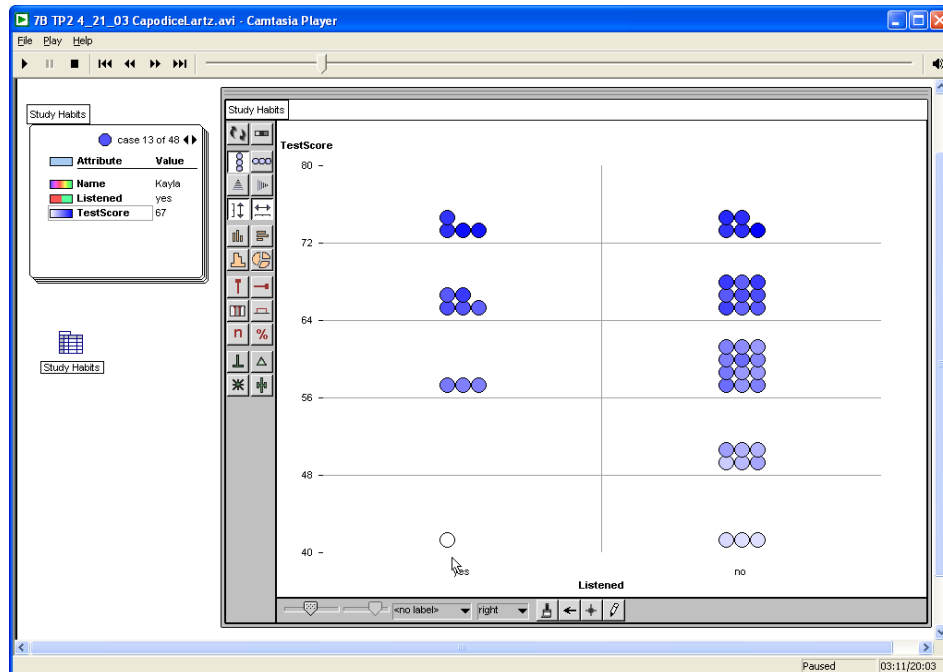


Figure 26. Side-by-Side Vertical Dot Plot

Episode 1 illustrates how Camille and Larissa analyzed the dot plot they constructed.

### Episode 1

- 18 C: Yeah. So like that's an A, that's a B, that's a C, that's a D, that's an F.  
 19 L: That's just bad.  
 20 C: Well okay.  
 21 L: Well...  
 22 C: It looks like people that studied classical music did very well.  
 23 L: It looks like they did better.  
 24 C: Pretty well on the test. They got a C or better.  
 25 L: Except for that one but he is an exception.  
 26 C: Yeah  
 27 L: Sean's a dud. He's an outlier.  
 28 C: Maybe Sean didn't listen to classical music and just put an x by his name anyway. Okay so...  
 29 C: I think it helped a little.  
 30 L: I think maybe because but you see that there is not that many people that listened to classical music and look at all these people that did listen to classical music and did a good job.  
 31 C: You're right.  
 32 L: Maybe they are just smart and don't need classical music but I don't know.

In their analysis of Figure 26, the pair compared student test scores to grades and

commented that some of the grades were “bad” (lines 18 – 19). Then without any further analysis the pair hypothesized that the students who listened to classical music did better (lines 22 – 23). They justified this hypothesis by saying that the students that listened to classical music “got a C or better” (line 24). Camilla and Larissa noted that their hypothesis was true, except the student who scored the lowest, Sean, who was a “dud” and also an outlier (line 27). They offered a reason for Sean’s low score: “Maybe Sean didn't listen to classical music and just put an x by his name anyway” (line 28).

Following their discussion about the student who scored the lowest, Camille hypothesized that “I think it helped a little” (line 29). They posited a reason why they believed the students who listened to classical music scored better because “Maybe they are just smart and don’t need classical music...” (line 32). Larissa also added insight to the Camille’s hypothesis, when she explained that “I think maybe [listening to classical music helps] because but you see that there is not that many people that listened to classical music and look at all these people that did listen to classical music and did a good job” (line 30). She recognized that the number of students in the two groups was different and that most of the students that listened to classical music had high scores.

They pair continued to analyze the data displayed in Figure 26. Episode 2 illustrates Camille and Larissa’s continued analysis.

#### Episode 2

- 35 C: The ones with the x, they’re all above 60. (*They are looking at the data table.*)  
36 L: Almost  
37 C: But except for Sean -- but he's a loser.  
38 L: 61, 60, 61, yeah they all got above 60.  
39 C: And that's a B.  
40 L: They all got above 60.  
41 C: That's a B.  
42 L: Yeah

- 43 C: But like those are wide spread.  
 44 L: Yeah those are, but you see, most of them got, maybe 50, they all got pretty close to a 62, and there is a lot of them there.  
 45 C: Yeah, But like the majority for these are up and the majority is down here.  
 46 L: Yeah, that's true, wait there's one think. Okay.  
 47 C: You know what I mean.

The pair noticed that all of the students that listened to classical music scored above 60 (line 35, 38, 40), that is, except for the student Sean, who scored the lowest and who they referred to as a loser (line 37). They observed that there was a gap between Sean's score and the score for the students who listened to classical music in the bin ranging from 56 to 64 points (lines 43 - 44). Camille referred to these being widespread. Camille recognized that the majority of scores for students that listened to classical music were up high on the graph, and the majority of scores for students that did not listen to classical music were down lower (line 45).

After Episode 2, the pair hypothesized that "they did a little better when they listened to classical music," saying "I think it helps a little bit" (episode 3, lines 49, 54). Episode 3 illustrates Camille and Larissa's discussion about their hypotheses and final conclusion to the questions.

### Episode 3

- 48 L: I think maybe it sort of helps.  
 49 C: Yeah, it helps a little bit.  
 50 L: Yeah it helps a little, maybe like not that much.  
 51 C: I am not going to start doing it, but...  
 52 L: Yeah okay. Um, so, for the first question. The test scores generally were about the same but a few people like there's they did a ...  
 53 C: There are exceptions.  
 54 L: They did a little better when they listened to music. Two...  
 55 C: I think it helps a little bit but...  
 56 L: Not enough to make a difference.  
 57 C: Don't go out of your way to do it.

58 L: I agree. I think maybe because but you see that there is not that many people that listened to classical music and look at all these people that did listen to classical music and did a good job.

Camille and Larissa both agreed that listening to classical music helped “somewhat.”

Camille said that “it helps a little bit” (line 49) and Larissa said “it helps a little, maybe like not that much” (line 50). They believed that listening to classical music did not make a big difference in test scores (line 56) and a student should not go out of their way to listen to it (line 57). Larissa reasoned that the students who listened to classical music did good on the test (the pair’s criterion for a good test score was 56), but there were just a many people that did not listen to classical music that did just as good (line 58).

**Characterization of Pair 7B TinkerPlots session.** Table 10 is a chronological listing of how pair 7B’s TinkerPlots session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

7B Critical/Creative Thinking	7B Statistical Thinking	Description
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation	Produces side-by-side vertical dot plot
Concretize		Discusses scores as grades
Concretize		Refers to bad scores
Hypothesis		“It looks like they did better.”
Concretize		Discusses scores as grades
Knowledge Building	Extreme/Outlier	Uses word “outlier”
Causal Relationship		“Maybe Sean didn’t listen to classical music and put and just put an x by his name anyway.”
Refinement of Hypothesis		“I think it helped a little.”
Knowledge Building	Different Group Sizes	“I think maybe because but you see that there is not that many people that listened to classical music and look at all these people that did listen to classical music and did a good job.”
Causal Relationship	Distribution Chunk	
Causal Relationship		“Maybe they are just smart and don’t need classical music.”

Causal Relationship Concretize	Outlier/Extreme	Refers to lowest score
		“But except for Sean – but he’s a loser.”
		Discusses scores as grades
	Spread	“But like those are wide spread.”
	Different Group Size	“They are got pretty close to 60 too, and there is a lot of them there.”
	Spread	
	Distribution Chunk	“But like the majority of these are up and the majority is down here.”
Conclusion		“I think may it sort of helps.” “It helps a little bit.”
	Different Group Size	Observes difference in data set size

*Table 10. Chronological List of Codes for Pair 7B TinkerPlots Session*

The following section is a characterization of Camille and Larissa’s critical, creative, and statistical thinking in TinkerPlots. The pair created one graph that was discussed and analyzed in regards to the question being asked – a variation of a dot plot. The statistical techniques the pair used in the creation and analysis of their representation were classifying, sequencing, extreme/outlier, spread, distribution chunk, and recognition of different group sizes. The critical and creative thinking elements that the pair used in the analysis of their graphs were hypothesizing, knowledge building, concretizing, and the identification of causal relationships.

Throughout the session Camille and Larissa’s statistical thinking can be characterized by their use of distribution chunks. They compared the placement of scores to one another, noticing that the majority of the scores for students that listened to classical music were up high and the majority of students that did not listen were down lower. They recognized that despite there being fewer students who listened to classical music, most of their scores were up higher on the graph – meaning that they scored higher than the students who did not listen to classical music.

The girls’ work on this task can be characterized by an analysis cycle that

included analyzing the graph they created, stating a hypothesis about which group of students did better on the Classical Music test, continuing their analysis, then refining their hypothesis. The pair hypothesized about which group of students did better on the test two times during their analysis of Figure 26. These hypotheses were related, and were only separated by one interchange concerning Sean, a student who listened to classical music, yet scored the lowest overall. The second hypothesis was slightly different in its analysis of the data than the first. The first hypothesis was “it looks like they did better” – referring to the students who took classical music. The second hypothesis changed to “I think it helped a little.” This subtle change in their hypothesis was made after their observation that the student who scored the lowest overall listened to classical music.

The pair never technically verified their hypotheses. After their first and second hypotheses they continued to “build knowledge” about the data through continued analysis and discussion. This knowledge building also helped them form a final conclusion. It was not clear why the pair continued analysis after their first hypothesis.

A characteristic of the pair’s work that was a result of their analysis cycle was the refinement of their original hypothesis. The pair concluded that listening to classical music “helps a little bit.” Larissa added that it this classical music’s effect was “not enough to make a difference” and Camille said “don’t go out of your way to do it.” The pair recognized that that overall the students who listened to classical music scored higher, reasoning that the people who listened did good, but there were just as many students that did not listen and did just as good on the test.

The pair’s analysis of the data on which this conclusion was based was another



characteristic of their thinking during the session. The pair examined clumps of students that scored above 56. This was a natural dividing point on the graph; 56 was the minimum score for the third highest bin. The two data sets were placed side-by-side on the graph making it easy to compare the number of student in each bin on a one-to-one basis. Camille and Larissa focused on the two clusters of students from both groups with the highest scores, partitioning them from the rest of the students. They commented that the scores of these two clumps of students were “about the same.” They failed to take all the scores for the students who did not listen to classical music into account. They did not compare the overall proportion of students within each group that scored high.

Another characteristic of Camille and Larissa’s work was adding context to the data. Concretizing is making a general idea specific by giving examples and applications which will make an abstraction meaningful. The pair converted raw scores to grades and categorized groupings of scores into grades A, B, C, D and F, with an F referred to a being bad. The topic of grades was present throughout the entire session. Since the pair was familiar with grades in their everyday school experience, this put the students test scores into a context they could better understand.

Camille and Larissa’s work can also be characterized by their identification of causal relationships. The pair often discussed Sean, a student who listened to classical music, yet scored the lowest out of all the students. They speculated about why Sean scored so low – saying he was an exception, and calling him a loser. They offered a reason for Sean’s low score: “Maybe Sean didn't listen to classical music and just put an x by his name anyway.”

The pair also attempted to explain why there were students that did not listen to

classical music, but had high scores: “Maybe they are just smart and don't need classical music but I don't know.” Their theory was that the students who listened to classical music did well on the test overall, but they would have done good whether they listened to classical music or not. They were just good students to start with is why they scored high on the test.

In summary, Camille and Larissa’s work in the TinkerPlots was characterized by their use of distribution chunks, their analysis cycle, the refinement of their hypotheses, their partial examination of the students who did not listen to classical music, their use of context, and the identification of causal relationships.

**Differences in thinking when working in the two mediums.** This section describes the differences in the pair’s thinking between the two mediums. A major difference in their work during the sessions was the how they represented the data. In paper and pencil the pair represented the data with fractions. In TinkerPlots they created one graph.

The analysis cycle during the session represented a major difference in the girl’s thinking. In paper and pencil, Camille and Larissa analyzed the data in one way and arrived at one response. They made no effort to continue their analysis or verify their response after reaching a conclusion. The pair spent most of the session constructing the fractions and very little time analyzing them.

In TinkerPlots the girls’ completed one analysis cycle. A result was the refinement of their first hypothesis “It looks like they did better” stating that “I think it helped a little.” This subtle change was made after their observation that the student who scored the lowest overall listened to classical music. The lowest score was highly visible

on the graph and the girls' felt a need to explain its placement. Their conclusion was more complex than their original hypothesis as it was modified to account for the outlier. It should be noted that the pair never technically verified their hypotheses. After their first and second hypotheses they continued to "build knowledge" about the data through continued analysis and discussion. This knowledge building also helped them form a final conclusion. It was not clear why the pair continued analysis after their first hypothesis.

Throughout the TinkerPlots session Camille and Larissa's statistical thinking can be characterized by their use of spread and distribution chunks. They observed gaps between scores and how tightly scores were grouped together. They compared the placement of scores to one another, noticing that the majority of the scores for students that listened to classical music were up high and the majority of students that did not listen were down lower. During their analysis they also examined clumps of students that scored above 60. This is similar to their analysis of the data in paper and pencil where they had a cut-off score and examined students who scored above 50. In paper and pencil the pair created fractions that represented this quantitatively, in TinkerPlots their means of comparing scores based on this cut-point was qualitative only.

A significant difference in Camille and Larissa's thinking in the two environments was how they used context in the analysis of the data. In TinkerPlots the pair made use of concretization and causal relationships, while in paper and pencil the pair only demonstrated concretization. In both environments the pair demonstrated five instances of concretizing encompassing the topics of grades and the low score. In addition, in TinkerPlots, Camille and Larissa identified causal relationships, often

speculating why Sean, a student who listened to classical music, scored so low. The pair also attempted to explain why there were students that did not listen to classical music, but had high scores.

There were differences in how the pair dealt with the difference in group size between the two environments. In both environments the pair recognized this difference. In TinkerPlots the pair overcame this difference by using distribution chunks to analyze the data. In paper and pencil, the pair chose to create fractions which made it possible to compare the data sets on an equal basis.

### **Pair 7C**

**Summary of Pair 7C paper and pencil session.** Pair 7C was comprised of two seventh-grade students, Lori and Nathan. The task they worked on in paper and pencil was called Drivers' Education (see Appendix C). In this task, participants are asked to determine whether students taking a drivers' education course performed better on their drivers' exam than those who did not. Data were presented to the students in tabular form and as a set of cards.

At the beginning of the session, Lori and Nathan organized the cards by dividing them into two groups: students who took Drivers' Education and students who did not take Drivers' Education. Using the top of the desk where they were seated, Lori and Nathan used writing pens and the cards to create a graphical representation of the data (Figure 27). The pair used the writing pens to create a grid with four quadrants. The vertical line of pens was the dividing line between the students that took Drivers' Education and those that did not. The score cards for those that took Drivers' Education were placed to the right side of the line and the score cards for those that did not were

placed to the left.

Lori and Nathan determined that it would help them analyze the data if the test scores were further separated. To do this, they used the pens to create a horizontal line that intersected the vertical line. To establish where the cutoff line would be in relation to the test scores, Lori reasoned that it should be the mid-point between the highest and lowest of all test scores. Since the highest test score was 70 and the lowest was 30, she concluded that the cutoff should be a test score of 50.



*Figure 27. Four Quadrant Table on Desktop*

Lori and Nathan then separated the cards according to their score, placing the cards with scores above 50 in the upper two quadrants and the cards with scores below 50 in the lower two quadrants. If a score was exactly 50 the pair would place that card aligned to the right or left of the horizontal divider. The scores above 50 were considered “good” and the scores below 50 were considered “bad.”

Episode 1 illustrates how Lori and Nathan analyzed the four quadrant graph they constructed.

#### Episode 1

- 29 N: So these guys are failing. (*Referring to the scores below 50*)  
30 L: So we can see if they had a lower or higher than the fifty. See?  
31 N: So 51, Amanda Buns. (*Referring to classmate.*)

- 32 L: Do you want to help me?  
33 N: Yea.  
34 L: Good job.  
35 N: Those people are failing oh my gosh. (*Referring to the scores below 50.*)  
36 L: The 50 would be right on the pen.  
37 N: No it wouldn't.  
38 L: Yes  
39 N: Then there would be only one failing person. I guess. You got more people that didn't get fifties.

During their discussion about the graph they had created, they engaged in the use of contextualization. For example, he referred to a card that had the name Amanda written on it as a classmate, "Amanda Buns" (line 31). Nathan considered a score below fifty to be failing and during episode 1 he made several references to the scores below 50 as failing (lines 25, 35, 39).

Lori and Nathan decided to transfer their desk-top representation consisting of cards and markers to paper and pencil so that they could record the number of students in each quadrant (Figure 28, recreation). The representation was the same as Figure 27 except that they wrote the number of students in each quadrant and the number of students who scored exactly 50 points out to the side of the horizontal marker. Furthermore, Lori and Nathan added labels and totaled the number of students in both the "Not Took" and "Took" columns.

	<b>Not Took</b>	<b>Took</b>	
	17	10	
50---3			50---2
	15	1	
	<b>Total = 35</b>	<b>Total = 13</b>	

*Figure 28. (Recreation) Four Quadrant Table on Paper*

Episode 2 illustrates how Lori and Nathan analyzed the four quadrant table they constructed on paper.

Episode 2

- 62 L: Anyways, they had more people, not take the test obviously.
- 63 N: Yeah there was 35 people that did not take the test and 13 people that did.
- 64 L: Um, so you can tell that the who did not, oh you can tell that the majority of people that did take the test got above 50.
- 65 N: And the majority of the people that didn't take the test got above 50.
- 66 L: It was still pretty much even, because 15 and 17 are still two away.
- 67 N: This is like the problem Monday. I don't even think it matters. What was that problem about?
- 68 L: Well the people that took Drivers' Education...
- 69 N: Well the people that took it did good. And the students who didn't take it still did good.
- 70 L: No they didn't!
- 71 N: Yeah they did!
- 72 L: They did average, because look they had like...
- 73 N: What is a 50 out of 70?
- 74 L: Nathan. They have 15 and 17 are only 2 away from each other.

The pair recognized that the majority of students in both groups scored above 50 points (lines 64 - 65). Nathan contended that this meant that both groups of students did "good" when he made the statement "The people that took it did good and the people who didn't take it still did good" (line 69). Lori disagreed that the people who did not

take Drivers' Education did good because there were roughly the same number of students that took Drivers' Education who scored above 50 compared to those who took Drivers' Education and scored below 50 (17 versus 15) (lines 66). Instead, she thought that the students who did not take Drivers' Education did "average" (line 72).

As the pair continued their analysis Nathan became confused about which quadrant the students who scored exactly 50 points should be placed. Lori responded by changing the horizontal "cutoff" to a test score of 49 points because there were no scores that were exactly 49 points. This, in turn, changed the number of students in the two upper quadrants of their graph (Figure 29).

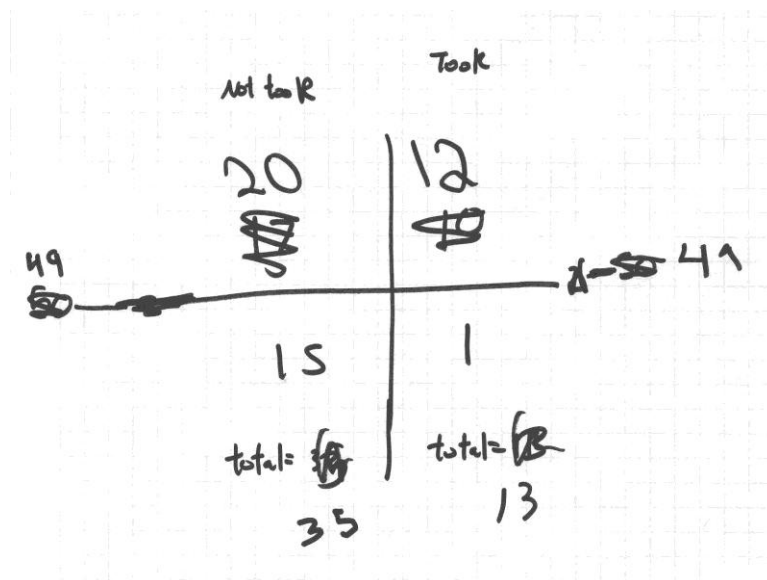


Figure 29. Modified Four Quadrant Table on Paper

Episode 3 illustrates how Lori and Nathan analyzed the modified four quadrant graph they constructed.

### Episode 3

- 83 N: The people who took it did better than the people who didn't take it.  
 84 L: Yeah.  
 85 N: The people who didn't take it still did good.



- 86 L: But they still had, like ok, think of it this way, the people who took it they only had one of them not get a very good score. But the people who didn't take Drivers' Education they had about half of them, or a little under half.
- 87 N: Three-fourths
- 88 L: Yeah like three-fourths not do that well. That means that people who took it were more successful.
- 89 N: 35 people didn't take it, out of the 35 um, like, 1/3 did bad and the other 2/3 did good.
- 90 L: I see what you are saying. Do you think the test scores – so then the test scores for the people who took driver's education were better than the people who did not take Drivers' Education.
- 91 N: Yeah.
- 92 L: Ok, what conclusion can you make about the effect of taking driver's education? It helps you.
- 93 N: Yeah.
- 94 L: Its better if you take it because you'll have a better...
- 95 N: It is like a study guide because then you know the answers.

Despite Lori's modification of the graph, Nathan continued to insist that the students who did not take Drivers' Education did "good" (line 85). Lori explained her hypothesis by using fractions to explain that the students who took Drivers' Education performed well (lines 86, 88). She explained that fewer students in proportion to the total number of students that took the course did poorly while a greater proportion of students who did not take the course did poorly. It appears that Lori analyzed the data multiplicatively describing that "a little under half" of the students who did not take Driver's Education scored below 49 (Line 86).

Nathan compared the fraction representing the number of students who did not take Driver's Education and who performed poorly ( $15/35$ ) with the fraction representing the number of students who did not take Driver's Education and performed well ( $20/35$ ). He estimated these fractions to be  $1/3$  and  $2/3$  (line 89). By looking at the fractions he saw that a greater proportion of students that took Drivers' Education did better on the test than students that did not take Drivers' Education. Nathan viewed the data additively

observing that 20 students that did not take Driver’s Education had scores over 49 and 15 had scores lower than 49.

Lori and Nathan determined that taking Drivers’ Education helped the students who took it (lines 92 – 93). Nathan even compared taking Drivers’ Education to a study guide when he said “It’s is like a study guide because then you like know the answers” (line 85).

**Characterization of Pair 7C paper and pencil session.** Table 11 is a chronological listing of how pair 7C’s session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

7C Critical/Creative Thinking	7C Statistical Thinking	Description
	Extreme/Outlier	Uses word “outlier”
	Classifying	Splits data cards into two piles
	Sequence	Orders data cards
	Extreme/Outlier	Identifies highest and lowest score
	Average	Finds midpoint of high and low score
Concretize		Discusses students that are failing
Concretize		Compares students to fellow classmates
	Sequence	Orders data cards
	Representation 1	Produces four quadrant graph on desktop
Concretize		Discuss how driving is similar to riding a Go Kart
	Modifying Representation	Modifies four quadrant graph on desktop
	Representation 2	Produces four quadrant graph on paper
Concretize		Compares scores to grades
	Different Group Size	Notices difference in data set size
	Distribution Chunk	“The majority of people that did take the test got above 50.”
Hypothesis		Nathan: “Well the people that took it did good. And the students who didn’t take it still did good.”
		Lori: “No they didn’t!” “They did average.”
	Spread	Discuss how far scores are from each other
Verify	Modifying Representation	Changed horizontal axis to a score of 49

Re-statement of Hypothesis		Nathan: “The people who took it did better than the people who didn’t take it.” “The people who didn’t take it still did good.”
	Proportional Reasoning	Compares fractions of students scoring above and below 50
Conclusion		“The test scores for the people who took Drivers’ Ed were better than the people who did not take Drivers’ Ed.
Analogical Thinking		“It is like a study guide because then you know the answers.”

*Table 11. Chronological List of Codes for Pair 7C Paper and Pencil Session*

Lori and Nathan represented the data with a four quadrant table and analyzed the data by producing fractions based on it. During the session the pair used the following statistical thinking elements to create their representations and analyze the data: extreme/outlier, classifying, sequence, average, recognition of different group size, distribution chunk, and spread. The pair used the following critical and creative thinking elements to analyze the data: concretizing, hypothesizing, verifying, and analogical thinking.

Lori and Nathan’s work on this task can be characterized by an analysis cycle that included analyzing the graphs they created, stating a hypothesis about which group of students did better on the Driver’s Education test, and attempting to verify the hypothesis. This cycle was repeated two times during the session. What prompted these cycles was the pair’s difference in opinion about which group scored better on the test.

The pair’s work on this task can be characterized by their difference of opinion about which group scored better on the test. Lori and Nathan both recognized that the majority of students in both groups had test scores above 50. But, the pair had a difference of opinion about what it meant. Lori developed a hypothesis that the group who took drivers education did better overall, based on distribution chunks and

proportional reasoning. Lori analyzed the data multiplicatively describing that less than half of the students who did not take Driver's Education scored below 50. Nathan did not understand this line of reasoning, or the effect of the different group sizes. His partner, Lori tried to help him understand these concepts which subsequently led to verification of the hypothesis and modification of their graph.

To the contrary, Nathan stated "The people that took it did good and the people who didn't take it still did good." It was his opinion that because the majority of students in both groups of students scored above 50 points, both groups of students did "good." Nathan analyzed the data additively observing the number of students that scored above and below 50. Nathan placed both groups of students on an equal level since both groups had students that scored above 50. Unlike Lori, however, Nathan was unable to differentiate which group scored proportionately higher than the other.

Lori and Nathan's work on this task can be characterized by a method of analysis that I termed the "cut-off technique." For this technique, fractions were created based a "cut-off" for good. Then the number of good scores were tallied, and divided by the total number of students in the group. Pair 7C created the two fractions they analyzed based on the numbers in their four quadrant graph.

Using proportional reasoning to compare these fractions also characterized the pair's work in paper and pencil. Lori explained only one student that took Drivers' Education had a low score, while half of the students who did not take Drivers' Education had above a score of 50 and half of them below. This argument represented a breakthrough in Nathan's thinking. He estimated the fractions for the students who did and did not take Drivers' Education ( $15/35$  and  $20/35$ ) to  $1/3$  and  $2/3$ , which helped him realize

the proportion of students who took drivers education and did good was much higher than this. After their analysis of the fractions, the pair agreed on their conclusion that taking Drivers' Education helped.

The way that pair organized the data in four quadrant table enabled them to produce representations that highlighted the distribution of scores so that they could interpret the differences between the groups. Although they acknowledged that the groups were different sizes, it was really the distribution of scores (greater than or less than 50) that prompted Lori to introduce a multiplicative comparison. Another characteristic of the Lori and Nathan's work in the session was incorporation of context into their discussion in the form of contextualizing and analogical thinking. By contextualizing the data Lori and Nathan made the abstraction of the data more meaningful to them. The pair personalized the data by referring to one of the students who took Drivers' Education, Amanda, as a fellow classmate, "Amanda Buns." Also, Nathan made several references to the scores below 50 as "failing." Nathan also demonstrated analogical thinking by comparing taking Drivers' Education to a study guide: "It's is like a study guide because then you like know the answers."

In summary, Lori and Nathan's work in the paper and pencil environment was characterized by their analysis cycle, the partner's difference of opinions, the "cut-off" technique they used to represent the data, proportional reasoning, and adding context to the data.

**Summary of Pair 7C TinkerPlots session.** The second task Lori and Nathan completed required them to determine whether students studying for a mathematics test while listening to classical music performed better than those who did not (see Appendix

B). The data in TinkerPlots were presented to the participants in a random arrangement of 48 circles in differing shades of blue. Each circle represented a student and his or her score on the mathematics test. Each circle was filled with a shade of blue that reflected the student's test score. Darker shades of blue represented higher scores whereas lighter shades of blue represented lower scores. Refer to Appendix A for an image of the way the data were presented to the students.

Lori and Nathan began the session by creating a bar graph that was ordered by test score and colored to show whether or not the students listened to classical music (Figure 30). The students that listened to classical music were represented with the color green and the students that did not listen to classical music were represented with the color red.

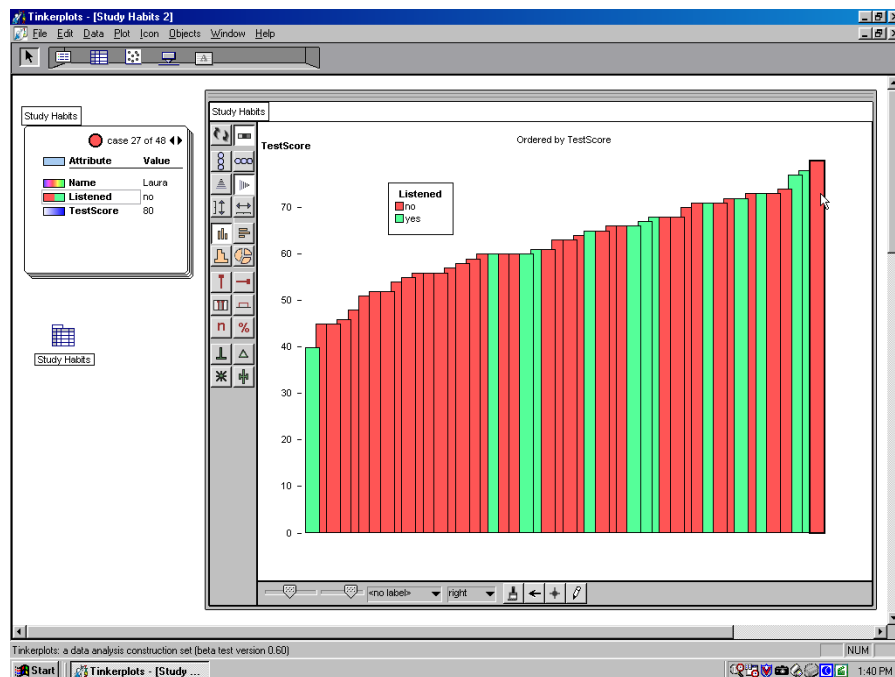


Figure 30. Bar Graph

Episode 1 illustrates how Lori and Nathan analyzed the bar graph they constructed.

### Episode 1

- 25 L: That person listened and their test score was like all the way down here.  
(Referring to Sean, the student who scored the lowest.)
- 26 N: Whoa, that person didn't listen and they still did awesome. (Referring to  
Laura, the student who scored the highest.)
- 27 L: Yea they did, it's green. That means, oh, that one, yea, was that Laura?
- 28 N: Well yea, Laura didn't listen.
- 29 L: And she still got a...
- 30 N: Sean did so bad.
- 31 L: 80 (Referring to Laura, the student who scored the highest.)

Their discussion regarding Figure 30 focused on the extreme scores. The pair observed that the student who scored the lowest, Sean, listened to classical music (line 25) and the student that scored the highest, Laura, did not listen to classical music (line 26). They referred to Sean's score as "bad" (line 30).

Lori and Nathan decided that Figure 30 was too complicated so they resolved to make a new one. They thought that a graph comprised of circles instead of bars would be easier to analyze. They created a new graph was split into five columns each containing an eight point range of scores (Figure 31). Individual student test scores were represented by circles and each column included both students who listened to classical music and those who did not listen. The students that listened to classical music were represented with the color green and the students that did not listen to classical music were represented with the color red.

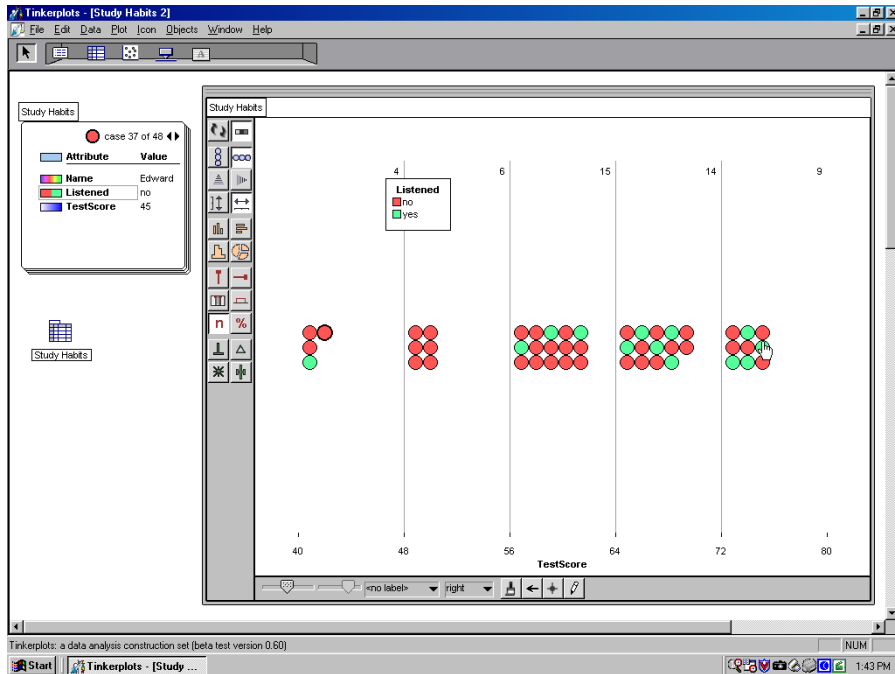


Figure 31. Five Column Dot Plot

Episode 2 illustrates how Lori and Nathan analyzed the five column dot plot they constructed.

Episode 2

- 57 L: Ok. So you can see that well a lot of the people that didn't, a lot of people didn't listen.
- 58 N: Yeah and they didn't do so well. Well a lot of people who listened also didn't do so well.
- 59 L: Yeah, see...
- 60 N: The majority of people that didn't listen got like Fs. That's weird.

In their analysis of Figure 31 the pair noticed that there were many students who did not listen to classical music (line 57), and these students did not do well on the test (line 58). They also observed there were many students who listened that didn't do so well (line 58). Nathan personalized this observation by explaining that "The majority of people that didn't listen got like Fs. That's weird" (line 60). After making this statement Nathan began to manipulate the settings of Figure 31. Once it became clear that Nathan would be



unable to restore the graph its original state the pair decided that they would create another graph.

The pair altered Figure 31 by combining the bins and placing the circles on a horizontal axis (Figure 32). The dot plot they created was ordered by test score and colored to show whether or not a student listened to classical music or not. The students that listened to classical music were represented with the color green and the students that did not listen to classical music were represented with the color red.

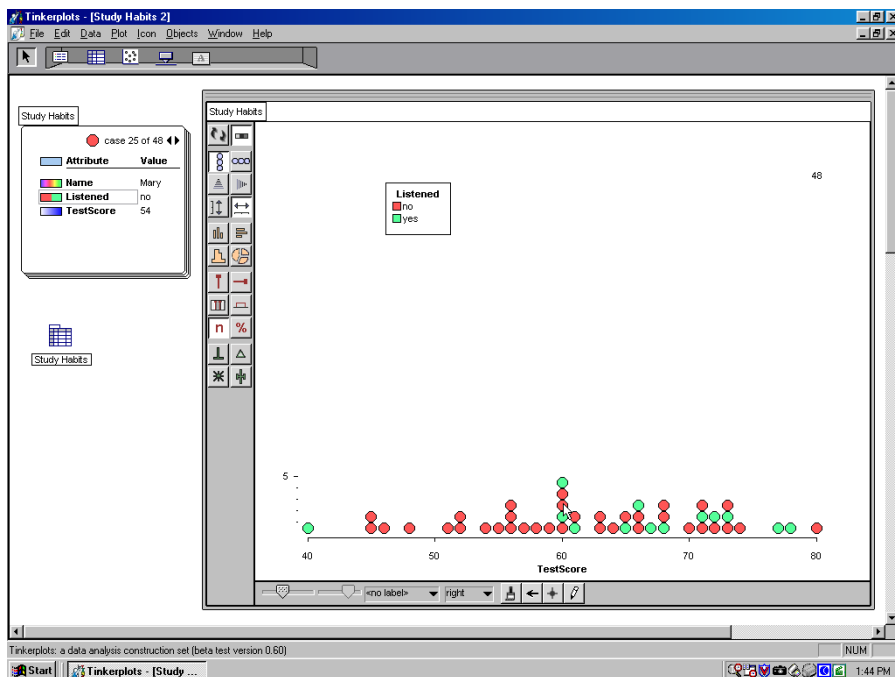


Figure 32. Dot Plot

Episode 3 illustrates how Lori and Nathan analyzed the dot plot they constructed.

Episode 3

- 82 N: Yeah. Or like, it was kind of like it was a distraction because if you like break it down to 60 out of 80 and then put 20 to each one. I guess that's good because they got Bs and didn't fail. That's good... got Cs.
- 83 L: And then the people who didn't listen to classical music were down to the left, but still there were some up.
- 84 N: Oh there were 48 kids in the class.
- 85 L: Yeah there were still some up here though, you see, that is the thing though.

- 86 N: I think like it was like good for some people and then it was more like a distraction for some people.
- 87 L: Yea.
- 88 N: Because they listened and that guy I don't know what was up with him. (*Referring to Sean, the lowest score.*)

In their analysis of Figure 32, Lori and Nathan observed that a lot of students that did not listen to classical were down towards the lower scores on the graph, but a lot of these students were also up towards the higher scores on the graph (lines 83, 85). Nathan focused on converting raw test scores to grades (line 82). He pondered the effects of listening to classical music while studying and said "I think like it was like good for some people and then it was more like a distraction for some people" (lines 82, 86). This statement was based on his observation that the student who scored the lowest, Sean, listened to classical music. By stating that listening to classical music was a "distraction for some people," Nathan was suggesting a reason why Sean scored so poorly.

After their discussion regarding Figure 32, the pair hypothesized that test score "depends on the person" (episode 4, line 95). The pair's hypotheses and final conclusion is chronicled in episode 4.

#### Episode 4

- 95 L: What conclusion can we make about the effect of classical music while studying?
- 96 N: Don't do it, I guess.
- 97 L: I think you can, it just depends on the person.
- 98 N: Personally I think it is Ok because it is kind of relaxing.
- 99 L: Nathan! The graph, think about the graph.
- 100 N: Oh ok.
- 101 L: I think it that it depends on the person because some of the people that listened to it didn't do well at all but some of them did really well.
- 102 N: Yeah.
- 103 L: And some of the people who didn't listen to it did really well and some of them who did, didn't listen to it did really poorly so.
- 104 N: I don't know it is kind of weird. It kind of played a part. It didn't really because people who did listen to it probably didn't study and people who did

- listen to it probably didn't study. But still I don't know. It's weird because I don't think classical music had an effect on the people.
- 105 L: I don't think it did either.
- 106 N: I think it was just them studying by themselves and the classical music. Because for them it helped, for them it was more like a distraction.
- 107 L: I think it is just mainly like your preference. And how well you can do on the test and how well you can study. I don't think it has much do with what kind of music they're listening to.
- 108 N: Because if it was rap...
- 109 L: Nathan!
- 110 N: No I mean, any other kind of music like rap and rock that would have been really distracting. But classical is not as distracting but...
- 111 L: Because people don't know the words.
- 112 N: Classical is more like... What was I going to say? Yea classical is more like, it's not supposed to be distracting but anything that, if you're studying and even picking up a crumb off of the table, you are going to pick up the crumb instead of studying.

Nathan hypothesized that test scores “depend on the person” (line 95). The pair reasoned that test scores were dependent on the individual student; how well the student studied or the type of music that they listened to (lines 104, 106, 107 – 112). Lori stated that “I think it is just mainly like your preference. And how well you can do on the test and how well you can study” (line 107). The pair concluded that classical music did not have an effect on test scores (line 104, 105).

Nathan provided additional rationale for their conclusion. He said that if the students had listened to rap, that would have really been distracting, but classical music is supposed to be relaxing, but even so, classical music can be equally distracting (lines 110, 112). He provided an analogy to describe how listening to classical music can be distracting "Yea classical is more like, it's not supposed to be distracting but anything that, if you're studying and even picking up a crumb off of the table, you are going to pick up the crumb instead of studying" (line 112).

**Characterization of Pair 7C TinkerPlots session.** Table 12 is a chronological listing of how pair 7C’s session was coded for critical, creative and statistical thinking. Included is a brief description of how the pair used each code. This listing is more comprehensive than what is portrayed in the summary above. It includes all the coded data, not just what is evident in the episodes presented.

7C Critical/Creative Thinking	7C Statistical Thinking	Description
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 1	Produces bar graph
	Extreme/Outlier	Discuss lowest and highest scores
Concretize		Refers to scores as bad
	Modifying Representation	Modifies bar graph
Concretize		Refers to scores as bad
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 2	Produces five column circle graph
	Different Group Size	Observes difference in data set size
Analysis Based Observation		“Well a lot of people listened and didn’t do so well.”
Knowledge Building Concretize		Refers to scores as Fs
	Range	Observes students scored from 56 to 64
	Modifying Representation	Modifies five column graph
	Classify	Splits data into two columns
	Sequence	Orders data from least to greatest
	Representation 3	Produces dot plot
	Extreme/Outlier	Uses word “outlier”
	Distribution Chunk	“Most of the reds are down there.” “Most of the people that didn’t listen are down there.”
		“It was kind of like a distraction.”
		Discusses scores as grades
	Distribution Chunk	“And then the people who didn’t listen to classical music were down to the left, but still there were some up.”
Analysis Based Observation		“I think it was good for some people and then it was more like a distraction for some people.”
Knowledge Building Causal Relationship		
	Extreme/Outlier	Discuss low score
	Different Group Size	Observes difference in data set size
	Range	Draws a box around scores from 45 through 59
	Distribution Chunk	
Hypothesis		Nathan: “Don’t I guess.”
Hypothesis		Lori: “I think it is okay because it is kind of relaxing.”
Causal Relationship		

Hypothesis		Lori: "I think it that it depends on the person because some of the people that listened to it didn't do well at all but some of them did really well."
	Distribution Chunk	Lori: And some of the people who didn't listen to it did really well and some of them who did, didn't listen to it did really poorly so."
Conclusion		Nathan "It kind of played a part. It didn't really because people who did listen to it probably didn't study and people who did listen to it probably didn't study. But still I don't know. It's weird because I don't think classical music had an effect on the people."
Causal Relationship		
Conclusion		Lori: "I don't think it did either."
Causal Relationship		Nathan: "I think it was just them studying by themselves and the classical music. Because for them it helped, for them it was more like a distraction."
Causal Relationship		Lori: "I think it is just mainly like your preference. And how well you can do on the test and how well you can study."
Causal Relationship		Nathan: "Because if it was rap..."
Analogical Relationship		Nathan: "Classical is more like... What was I going to say? Yea classical is more like, it's not supposed to be distracting but anything that, if you're studying and even picking up a crumb off of the table, you are going to pick up the crumb instead of studying so."

*Table 12. Chronological List of Codes for Pair 7C TinkerPlots Session*

Lori and Nathan analyzed three graphs during the session. The statistical techniques the pair used in the creation and analysis of their representations were classifying, sequencing, modifying, extreme/outlier, distribution chunk, and recognition of different group sizes. The critical and creative thinking elements that the pair used in the analysis of their graphs were hypothesizing, knowledge building, concretizing, and the identification of causal relationships.

The critical and creative thinking of this pair can be best characterized by their cycles of analysis. During the analysis, the pair made several statements that related to

answering the question being asked, however, they were not in direct response to the question, hence not meeting the criteria for a hypothesis. I have coined these instances as “analysis-based observations.” For example, while analyzing Figure 31, Lori said “Yeah and they didn’t do so well. Well a lot of people who listened also didn’t do so well.” During their analysis of Figure 32 Nathan said “I think like it was like good for some people and then it was more like a distraction for some people.” These statements were valuable to helping Lori and Nathan understand the data. They were analytical in nature but did not present a testable explanation.

The pair continued to analyze and discuss the data, but since the analysis-based observations were not judgmental, they did not try to prove or verify them. They simply continued to build their knowledge about the data set. Their analysis-based observations demonstrated the pair recognized that most of the students who did not listen to classical music scored low and most of the students that did listen to classical music scored high.

Throughout the session the Lori and Nathan’s statistical thinking can be characterized by their use of distribution chunks. These clusters of scores were subsets within the distribution of students who scored the best and students who scored the worst. Their analyses were based on visual inspection of the graphs and qualitative comparisons. For example, regarding Figure 32, Lori described the placement of the red circles compared to the green circles. She observed that most of the reds were on the left side of the graph: “And then you can see that most of the reds are down there. Most of the people that didn’t listen are down there.” This was the side of the graph with the lowest scores. Lori was describing a distribution chunk, that is, the placement of a clump of scores within the distribution.

Another statistical aspect of the data that Lori and Nathan focused on was extreme scores. They discussed extreme scores during their analysis of all three graphs. This was significant because it influenced their hypotheses and final conclusion. Lori and Nathan concluded that classical music did not have an effect on test scores. The pair's rationale was that test scores were dependent on the individual student. Nathan argued that classical music was distraction for some students. This was based on his observation that the student who scored the lowest, Sean, listened to classical music. Lori argued that test scores were based on how well an individual studied since some students didn't study did really well and others that listened to it did poorly. Her comment about students who did listen and did poorly was a direct reference to Sean.

The pair's work on this task can also be characterized by the identification of causal relationships. Lori and Nathan also added context to the data by suggesting causes of an event or situation. These suggested causes were the pair's way of justifying their conclusion that listening to classical music did not have an effect on test scores. For example, Nathan stated that the students who listened to it probably didn't study, and Lori added that your score depended on how well you can study. These statements were the pair's way of rationalizing their conclusion.

In summary, Lori and Nathan's work in the TinkerPlots can be characterized by their use of the analysis cycle, distribution chunks, extreme scores, and the identification of causal relationships. A significant characteristic of the pair's analysis is that they never quantified the data. Rather, their analyses were based on a visual inspection of the graphs and qualitative comparisons.

**Differences in thinking when working in the two mediums.** This section describes the differences in the pair's thinking between the two mediums. In paper and pencil the pair represented the data using a table and in TinkerPlots they created three graphs.

In both environments the pair rotated through the analysis cycle several times. In paper and pencil the pair demonstrated the thinking elements hypothesizing and verifying. In TinkerPlots, the analysis cycle was somewhat different. During these cycles the pair used analysis-based observations and knowledge building in lieu of hypothesizing and verifying. They continued to analyze and discuss the data, but since the analysis-based observations were not judgmental, they were not trying to prove, or verify them. The reason why they continued to analyze the data was unclear. TinkerPlots was a much more dynamic environment than paper and pencil, and freed them to "play around" with the data. Little effort was required to create and manipulate a graph in TinkerPlots. Manipulation of the data was easy to accomplish, which in turn, prolonged their analysis. In paper and pencil what prompted these cycles was the pair's difference in opinion about which group scored better on the test.

There was a difference in the pair's method of analysis in the two environments. In paper and pencil the pair quantified the data; each partner analyzed the data differently. Lori analyzed the data multiplicatively, for example, describing that less than half of the students who did not take Driver's Education scored below 50. Nathan analyzed the data additively observing the number of students that scored above and below 50. He placed both groups of students on an equal level since both groups had students that had scores above 50. Unlike Lori, however, Nathan was unable to



differentiate which group scored proportionately higher than the other.

The pair's four quadrant table in paper and pencil enabled them to produce representations that highlighted the distribution of scores, allowing them to interpret the differences between the groups. It was this organization of the data that prompted Lori to introduce a multiplicative comparison.

In TinkerPlots, Lori and Nathan's analysis of the data can be characterized by their use of distribution chunks, that is, the placement of clumps of scores within the distribution. In TinkerPlots the pair never quantified the data like they did in paper and pencil. Rather, their analyses were always based on a visual inspection of the graphs and qualitative comparisons.

In TinkerPlots, another statistical aspect of the data that Lori and Nathan focused on was extreme scores, discussing extreme scores during their analysis of all three graphs. This was significant because it influenced their hypotheses and final conclusion. In TinkerPlots the outlier was highly visible in all the graphs the pair created. Lori and Nathan concluded that classical music did not have an effect on test scores. The pair's rationale was that test scores were dependent on the individual student based on their observation that Sean, a student that listened to classical music scored very low.

This was not the case in paper and pencil. Extreme scores were not visible in the graph they created and did not play a role their analysis of the data. However, the extreme scores did play a role in the creation of their representation. To establish the cutoff line for "good" scores, they used the mid-point between the highest and lowest of all test scores.

In both mediums the pair incorporated a lot of context into their discussion. Lori

and Nathan engaged in concretizing five times during the paper and pencil session and seven times during the TinkerPlots session. A major theme that was contextualized was converting raw scores to grades. In TinkerPlots the pair's work can also be characterized by the identification of causal relationships. This was the pair's way of justifying their conclusion that listening to classical music did not have an effect on test scores. In paper and pencil the pair did not identify any causal relationships.

### **Findings of Cross-case Analysis**

The second major section of this chapter presents findings of my cross-case analysis. In this section I report my findings on the cross-case analysis of the paper and pencil task and the cross-case analysis of the TinkerPlots task. During Analysis Phase 3 I found patterns, themes, and trends within each environment. Relationships between the codes were found and categories emerged. I became aware of broader patterns in the data.

The broad themes that were identified across both tasks were the same: representation, context, the analysis cycle, method of analysis, and difference in group size. The commonalities and differences in the way pairs approached the tasks and their understanding of critical, creative, and statistical thinking in these areas are described below.

#### **Cross-case Analysis of the Paper and Pencil Task**

**Representation.** While working in paper and pencil the pairs created a variety of representations to describe the data – all of which were unique. In paper and pencil, no pair created more than one representation of the data. Fractions were the most common type of representation in paper and pencil. All four pairs that represented the data with fractions used some form of the “cut-off technique” to create them. Creating the fractions

was a very time intensive process leaving little time for analysis. These pairs did not generate their hypothesis until after the fractions were constructed near the end of the session.

A characteristic common among three of the pairs representations was their level of accuracy. Two of the pairs in paper and pencil created representations with errors in them, and one pair's representation was accurate, but not representative of the entire data set.

**Context.** During the paper and pencil sessions many of the pairs added context to the data. Concretizing was demonstrated by four of the six pairs. A major theme in concretization was converting scores to letter grades then discussing whether the scores were good or bad. Since the participants were familiar with grades in their everyday school experience, this put the students test scores into a context they could better understand.

**The analysis cycle.** In paper and pencil, four of the six pairs demonstrated an analysis cycle involving analyzing the representation they created, stating a hypothesis about which group of students did better, and attempting to verify the hypothesis. Two of the four pairs completed multiple rounds of the analysis cycle while two pairs completed a single cycle. The remaining two pairs analyzed the data in one way and arrived at one response, making no effort to continue their analysis or verify their response after reaching a conclusion.

It should be noted that for the two pairs completing one round, their analysis cycle was characterized as limited. These pairs' verification was more of an afterthought as it was not used to obtain more evidence to support their conclusion.

**Difference in group size.** In paper and pencil, most of the pairs recognized the difference in group size, and their method of analysis took this into account. The difference in group size was dealt with by the type of representation. Five of the pairs lumped the data for each group of students together into an aggregate score making it possible to compare the two groups of students on an equal basis. However, once their representation was created, the difference in group size never played a part in their analysis.

**Method of analysis.** In paper and pencil five of the six pairs quantified the data. The common way of doing this was creating fractions to represent the data. The pairs lumped all the data into two fractions and then compared their size. Only two pairs created graphical representations of the data. One of these pairs analyzed the data using fractions found within the graph. The other pair that created a graph identified patterns within the data using modal clumps and distribution chunks. The only two pairs in paper and pencil that used visual techniques to analyze the data and make qualitative comparisons were also the only two pairs that created graphical representations of the data.

### **Cross-case Analysis of the TinkerPlots Task**

**Representation.** In TinkerPlots there were many similarities between their representations that the pairs created. In TinkerPlots all of the graphs were variations of bar graphs, circle graphs, and dot plots. Many of the pairs created the same graphs, and sometimes a graph was created, then modified into another graph, and then changed back to the original. In TinkerPlots the pairs had the opportunity to see the data represented in many arrangements.

In TinkerPlots, the number of graphs that each pair created and then analyzed ranged from one to five. Only one pair created a single graph. The graphs were created and modified many times with ease. While some of the modifications were meticulously calculated to achieve specific changes to the graph, others were made with no specific purpose in mind. Due to the functionality of TinkerPlots, all of the graphs were complete and representative of the data.

**Context.** During the solution of the TinkerPlots task all six pairs concretized and identified causal relationships. The major aspect of the data that was concretized was converting raw scores to grades and discussing whether the grades were good or bad. And, also then comparing the grades between the students who took Driver's Education and those who did not to determine who got the better grades. Since the participants were familiar with grades in their everyday school experience, this put the students' test scores into a context they could better understand.

A major theme in the identification of causal relationships in TinkerPlots was the lowest overall score belonging to Sean, a student who listened to classical music. The participants tried to explain why Sean scored lower than any of the other students. With their explanations, the pairs speculated about why Sean did not fit the overall pattern and trend of other students who listened to classical music and did well.

Instances of concretizing and causal relationships were interspersed throughout the sessions. They occurred during the statistical analysis and alongside statistical thinking elements. However, there was a high incidence of causal relationships towards the end of the sessions that accompanied the participant's conclusions. These causal relationships were the pairs' way of justifying their conclusion that listening to classical

music did not have an effect on test scores, despite this group of students containing the outlier.

**The analysis cycle.** In TinkerPlots, all six pairs demonstrated an analysis cycle involving analyzing the representation they created, stating a hypothesis about which group of students did better, and attempting to verify the hypothesis. In TinkerPlots, four of the six pairs completed multiple rounds of the analysis cycle while the other two pairs completed a single cycle.

An outcome of the analysis cycles was the refinement of a pair's original hypothesis. As a pair continued to analyze the data, their hypotheses changed slightly with each cycle, adding to its complexity. Many caveats, explanations, and rationalizations were added to a pair's conclusions.

**Method of analysis.** In TinkerPlots the pairs' analyses were always based on a visual inspection of the graphs and qualitative comparisons. They were able to visualize the entire range of the data arranged from least to greatest, and how the scores of both groups of students were placed throughout. The pairs were able to perform a side-by-side comparison of the two groups of students. These visualizations helped the pairs analyze the placement of the test scores within the distribution and in relation to each other.

Each of the pairs made observations regarding the placement of the two groups of students in relation to one another on the graph. These insights were a major focus in their analysis of the data. Three related statistical thinking elements demonstrated throughout the TinkerPlots sessions that relied on visual inspection were spread, modal clumps, and distribution chunks. All of the pairs in TinkerPlots recognized and analyzed distribution chunks within the data.

In TinkerPlots, the extreme scores played a significant role in the hypotheses and final conclusions. The extreme low score (which was also an outlier) was frequently paired with the identification of causal relationships. The pairs could see the outlier and felt a need to explain it. The outlier was easily recognizable and prominent due to the contrasting colors and sequencing of the data from least to greatest.

**Difference in group size.** In TinkerPlots, the difference in group size was dealt with using visual analysis techniques. All of the pairs in TinkerPlots recognized the difference in group size, making statements regarding it. This awareness affected all of the pairs' analysis of the data. The pairs could see that the number of students who listened to classical music was different than the students that did not listen to classical music. The difference in group sizes was easily recognizable in the visual provided by each of the graphs.

### **Summary of Findings**

In this section I have presented findings of the cross-case analysis of the paper and pencil task and the cross-case analysis of the TinkerPlots task. During Analysis Phase 3, I found patterns, themes, and trends within each environment. Relationships between the codes were found and categories emerged. I became aware of broader patterns in the data. General findings across all pairs in each environment were identified. These general findings are reported on below.

General findings across all pairs were identified for the paper and pencil task. All the pairs spent more time on constructing representations than they did interpreting or analyzing the representations they had created, and every pair created a single representation. A general finding regarding context across all pairs was the absence of the

identification of causal relationships; only one pair identified a single causal relationship. In their analysis, all pairs quantified the data and the difference in group size was dealt with by the type of representation they chose.

General findings across all pairs were identified for the TinkerPlots task. The pairs created many representations that were complete and representative of the data, modifying them with ease. All the pairs added context to the data, including both concretization and causal relationships. Every pair demonstrated an analysis cycle involving analyzing the representation they created, stating a hypothesis about which group of students did better, and attempting to verify the hypothesis. The pairs' analysis was based on visual inspection of the graphs and qualitative comparisons. Each of the pairs made observations regarding the placement of the two groups of students in relation to one another on the graph using modal clumps and distribution chunks. All pairs were aware of the difference in group sizes and used visual analysis techniques to deal with it.

The research questions will be answered in the next chapter by merging the findings of the cross-case analysis and interpreting them with regard to the related research literature and the theoretical perspectives that guided the study.



## CHAPTER V

### CONCLUSIONS

This study investigated and characterized the types of higher order thinking and statistical thinking that students exhibited in technological and non-technological environments. In this chapter, I discuss my findings as they relate to the research questions, describe the implications for technology and instruction, limitations and constraints, and recommendations for future research.

#### **Overview of Study**

The purpose of this study was to gain a better understanding of how students think in a technological environment. This was accomplished by exploring the differences in the thinking of students while they worked in a technological environment and comparing this to their work in a paper and pencil environment. The software program, TinkerPlots, a construction tool that middle school students use for data analysis, was the technological environment used in this study. In both environments, types of critical, creative, and statistical thinking were characterized. The focus was on how students interpreted and analyzed data, and how those processes appeared to be affected by the two mediums, helped answer the two research questions in this study:

1. How does the critical and creative thinking of middle school students using the software program TinkerPlots differ from their thinking while using a traditional paper and pencil format?

2. How does the statistical thinking of middle school students using the software program TinkerPlots differ from their statistical thinking while using a traditional paper and pencil format?

This study involved the interplay of two theoretical perspectives. First, was the use of the technology acting as an intellectual partner with the user, lending itself to the support of cognitive processes. Second, was the concept of a computer functioning as a Mindtool, that is, a computer application that engages students in critical thinking. This theoretical perspective supports technology acting as an intellectual partner because Mindtools are computer-based learning environments that are designed to provide such partnerships (Jonassen, 2000).

The population of this study consisted of students in grades 6 and 7. Students' thinking was examined as they engaged in two tasks. One task was worked using the TinkerPlots software and the other task was worked using paper and pencil. The tasks were designed to elicit critical, creative, and statistical thinking. These tasks engaged students in comparing distributions and areas pertinent to comparing distributions including measures of central tendency, measures of variation, and the creation of representations.

This study utilized multiple sources of data: (a) video and audio tapes of students' problem-solving sessions, (b) recordings of computer actions and movements, (c) student artifacts, and (d) researcher observations and field notes.

Three phases of analysis were conducted. During the first phase, I developed a coding framework to adequately characterize the data. In the second phase of analysis, characterizations were developed for each session and differences in thinking were

compiled for each pair across both mediums. For the third phase of analysis, cross-case analysis summaries were developed for each environment.

Student thinking in a technological environment is an emerging field of study and there are limitations to the existing research. This study will help to fill the gaps in the research on statistics and technology. To address the limitations of the current research, this study will utilize the most up-to-date computer software package to teach statistics that was developed with the intent of being used with constructivist mode of learning. It will also be one of only a few studies where computer technologies are the central component.

### **Answers to Research Questions**

In this section, the research questions will be answered by merging the findings of the cross-case analysis for each environment and interpreting them with regard to the related research literature. The broad themes that were identified as common across environments will be used to organize the findings: representation, context, the analysis cycle, method of analysis, and difference in group size.

#### **Research Question One**

For research question one, the themes of context and the data analysis cycle were identified as differences in critical and creative thinking across the environments. The differences in the way pairs approached these themes in relation to critical and creative thinking are described below.

**Context.** Overall, the pairs added more context to the data in their TinkerPlots sessions. During the solution of the TinkerPlots tasks, all six pairs concretized and conjectured about causal relationships. Concretizing was demonstrated by four of the six

pairs in paper and pencil, but only one pair conjectured about a single causal relationship. So whereas, concretization was a similarity between the two environments, conjecturing about causal relationships was a difference.

In both environments, the major theme in concretization was converting scores to letter grades, then discussing whether the scores were good or bad. Since the participants were familiar with grades in their everyday school experience, this put the students' test scores into a context they could easily relate to.

A major theme in the conjecture of causal relationships in TinkerPlots was the lowest overall score, which belonged to Sean, a student who listened to classical music. The participants tried to explain why Sean scored lower than any of the other students. With their explanations, the pairs speculated about why Sean did not fit the overall pattern and trend of other students who listened to classical music and did well.

Instances of concretizing and conjectures about causal relationships were interspersed throughout the sessions. They occurred during the statistical analysis and alongside statistical thinking elements. However, there was a high incidence of conjectures about causal relationships towards the end of the sessions that accompanied the participant's conclusions. These conjectures about causal relationships were the pairs' way of justifying their conclusion that listening to classical music did have an effect on test scores, despite this group of students containing the outlier.

***Relationship to related literature.*** Research has shown the importance of context knowledge in student thinking (Gal, 2002; Moore, 1990; Papanastasiou & Melietiou-Mavrotheris, 2008; Pfannkuch & Rubick, 2002; Gal, 2004; Moore, 1990). Shaughnessy, Ciancetta, Best, and Canada (2004) found that about one-third of students included personal contexts

and past experiences in their responses on tasks comparing distributions. In my research, more than one-third of participants in both environments included context in the solution of the tasks, and in TinkerPlots, all participants included context in the solution of the task.

In the TinkerPlots environment of my study, I found that conjecturing about causal relationships was the pairs' way of justifying their conclusion that listening to classical music did not have an effect on test scores, despite this group of students containing the outlier. This resonates with the findings of Langrall, Nisbett, Mooney, and Janssen's (2011) study about the role of context expertise when comparing two data sets. They found that "students utilized context knowledge to provide additional justification or support for a claim they made in about 18% of all context knowledge occurrences" (p. 63).

Gal (2002, 2004) talked about "worry" questions that statistically literate adults should have in their mind to help them interpret data and claims presented in the media. These questions include ones such as "where did the data come from," "what kind of study was it," "was a sample used," and "how reliable or accurate were the instruments or measures?" (p. 16). These questions are important to be able to examine the reasonableness of claims regarding statistical information being communicated or displayed. When analyzing conclusions about information they have been presented with, adults must be concerned with the credibility of the evidence, the source of the information, and explore alternate explanations of those conclusions.

The responses to these questions support critical evaluation and lead to more informed interpretations and judgments of statistical information. Gal's model of statistical literacy was intended for consumers of statistics, however, it has been found to

be relevant to students in school settings (Langrall, Nisbet, Mooney, & Jansem, 2011). In my study, some students might have been taking a critical stance toward the data on the Classical Music task when they tried to interpret why Sean's test score was lower than the rest of the group. Other students, however, may have been trying to rationalize the data instead of questioning it.

Langrall, Nisbet, Mooney, and Jansem (2011) found that "students exhibited a tendency to rationalize the data; that is, they generated reasons why the data are the way they are" (p. 64). Although some students' explanations were based on their context knowledge and were statements of fact, at other times students did not have factual knowledge of the context and tried to create plausible explanation of the story behind the data. These findings relate to the conjectures about causal relationships that the pairs of students in my study posited while working in TinkerPlots. The students in my study did not have any expertise regarding the context of the tasks. Thus, they were not reporting a story behind the data; rather they were trying to create a plausible story by constructing explanatory theories to account for the data in a contextual way.

**The analysis cycle.** A general finding across all pairs in TinkerPlots was that they demonstrated an analysis cycle which involved analyzing a representation they created, stating a hypothesis about which group of students did better, then attempting to verify that hypothesis.

Hypothesizing and verifying were observed to occur in conjunction with one another. A pair would generate a hypothesis, then try to prove or show whether the hypothesis was correct. Sometimes this cycle was repeated multiple times. I called this pattern of hypothesizing, then attempting to verify the hypothesis the "analysis cycle."

In both environments, the act of hypothesizing used a combination of several thinking elements: inferring inductively, summarizing, and identifying main ideas. These elements were common across all pairs in both environments and were endemic to the tasks posed. In the process of hypothesizing, the students inferred deductively, that is, they “derived logical conclusions from accepted generalizations, truths, or principles” (Jonassen, 2000, p. 28). The hypotheses identified main ideas by finding the central meaning of tasks, and were summarizations by reducing the data into explanations of the information given to them.

Overall, it was observed that there was more analysis performed in TinkerPlots. Hypothesizing and verifying were demonstrated by all six pairs in TinkerPlots and only four pairs in paper and pencil. In TinkerPlots, four of the six pairs completed multiple rounds of the analysis cycle while only two pairs completed a single cycle. In paper and pencil, two of the four pairs completed multiple rounds of the analysis cycle while two pairs completed a single cycle. The remaining two pairs in paper and pencil analyzed the data in one way and arrived at one response, making no effort to continue their analysis or verify their response after reaching a conclusion.

In paper and pencil, there were several reasons why the pairs continued to analyze the data beyond their initial hypothesis including a lack of understanding of the effect of different group sizes and to check for accuracy of calculation. None of the pairs in TinkerPlots verified their hypotheses due to either of these reasons. In TinkerPlots, the reasons why the pairs continued the analysis process were varied. One pair verified their hypothesis due to being cautious and wanting to check their work. Another pair continued analysis due to disagreement over what had been hypothesized, while another pair’s

verification was triggered by the observation that there were disproportionate group sizes.

It was not clear why three of the pairs in TinkerPlots decided to verify their hypothesis and continue analysis. TinkerPlots was a much more dynamic environment than paper and pencil, and freed them to “play around” with the data – rapidly. Little effort was required to create and manipulate a graph in TinkerPlots. The pairs continued to analyze the data since manipulation of the data was easy to accomplish, which in turn, prolonged their analysis.

In TinkerPlots, an outcome of the analysis cycles was the refinement of a pair’s original hypothesis. As a pair continued to analyze the data, their hypotheses changed slightly with each cycle. In almost all cases the pairs’ conclusion was more complex in TinkerPlots. Many more caveats, explanations and rationalizations were added to the pairs’ conclusions in TinkerPlots. In paper and pencil, the pairs’ conclusions were typically blanket statements such as “taking Driver’s Education helped” or “the students who took Driver’s Education scored higher.” These were straightforward responses. However, in TinkerPlots the responses were richer and full of context. They often included casual relationships used for explanations of the outlier and justifications of their response.

***Relationship to related literature.*** Several studies using TinkerPlots have focused on its value as an analytical tool. Ben-Zvi (2006) found an advantage of using TinkerPlots was that students were able to use it successfully as an argumentative tool for presenting their ideas, as opposed to just a representational tool. Fitzallen and Watson (2011) reported participants in their study were able to quickly test multiple hypotheses when using TinkerPlots. The findings of my research are in agreement with both of these



studies.

## **Research Question Two**

For research question two, the themes of representation, method of analysis, and difference in group size were identified as differences in statistical thinking across the environments. The differences in the way pairs approached these themes in relation to statistical thinking are described below.

**Representation.** Representations in the two environments were very different. In TinkerPlots there were more similarities between their representations than there were in paper and pencil. In paper and pencil the pair's representations ranged from fractions to represent the two groups of students, to a back-to-back stemplot, to a four quadrant table. In TinkerPlots all of the graphs were variations of bar graphs, circle graphs, and dot plots. Many of the pairs created the strikingly similar graphs, and sometimes a graph was created, then modified into another graph, and then changed back to the original. In TinkerPlots the pairs had the opportunity to see the data represented in many more arrangements.

While working in paper and pencil the pairs created a variety of representations to describe the data – all of which were unique. In paper and pencil, no pair created more than one representation of the data and the most common type was fractions. All four pairs that represented the data with fractions used some form of the “cut-off technique” to create them. The fractions were created by determining a “cut-off” for good and bad scores. Then the number of good and/or bad scores were tallied and divided by the total number of students in the group. This technique was very creative; traditional mathematics and statistics textbooks do not introduce students to this method of analysis.

However, creating the fractions was a very time intensive process leaving little time for analysis. The pairs spent more time on constructing representations than they did interpreting or analyzing the representations they had created.

In TinkerPlots, the number of graphs that each pair created and then analyzed ranged from one to five. Only one pair created a single graph. In the TinkerPlots sessions, graphs were created and modified many times with ease. While some of the modifications were meticulously calculated to achieve specific changes to the graph, others were made with no specific purpose in mind.

Another major difference in the representations between the two environments was their level of accuracy. Two of the pairs in paper and pencil created representations with errors in them, and one pair's representation was accurate, but not representative of the entire data set. All of the graphs created in TinkerPlots were complete and representative of the data.

In Tinkerplots there were more similarities between students' representations than there were in paper and pencil. The data in TinkerPlots were presented to the participants in a random arrangement of 48 circles in differing shades of blue. Using the intuitive features of TinkerPlots, learners manipulated the data by rearranging it into different representations. They clicked on buttons in the left-hand column of the screen to order, sequence, merge, and separate the data. They pulled and dragged the pieces of data into new arrangements with differing scales and bin sizes. Although this led to the creation of a wide variety of representations, there were similarities between the graphs. In Tinkerplots all of the graphs were variations of bar graphs, circle graphs, and dot plots; with many of the pairs creating the same graph. The types of graphs created were limited

by the tools available in Tinkerplots.

In paper and pencil the pair's representations ranged from fractions to represent the two groups of students, to a back-to-back stem-and-leaf display, to a four quadrant graph. Creating fractions by summing totals was not possible in Tinkerplots. All of the representations in Tinkerplots, whether they displayed as separate data points or merged into one graph, still retained the all the individual pieces of data. In paper and pencil, the choice of representation was due in part to the way the data were presented to students, and also by the difference in group sizes. The pairs were given individual data cards for each student score. Summarizing the data into aggregate scores made the individual pieces of data easier to handle and also equalized the size of the data sets. While creating fractions was a time consuming process, the outcome did yield numeric representations that the students were able to analyze by comparing their size or converting to percentage. These techniques were appropriate for a middle school level of mathematical content knowledge.

*Relationship to related literature.* Studies have shown that students working in TinkerPlots create many different graphs quickly, analyzing them and then moving on to another graph (Hammerman & Rubin, 2003; Paparistodemou & Meletiou-Mavrotheris, 2008). In my study, I observed the same thing; students working in TinkerPlots created many graphs and modified them with ease.

In my study, four of the six pairs in paper and pencil created fractions to represent the data, while only two pairs constructed a graph. In TinkerPlots, all of the pairs constructed graphs to represent the data. These findings resonate with the work Meletiou-Mavrotheris, Paparistodemou, and Stylianou (2009) who conducted a study were

participants used both TinkerPlots and a paper-based environment to compare data sets. In the paper-based environment, participants represented the data with numerical calculations and did not create any graphs while in TinkerPlots, participants constructed graphs to represent the data.

In exploratory data analysis (Tukey, 1977; Hartwig & Dearling, 1979) visual representations are considered superior to numeric representations since they show the shape of the data. It is important to analyze the shape of the distribution before choosing statistical measures to describe its characteristics. Measures of location and spread were lost in the pair's creation of fractions. In addition, outliers and extreme values were combined into the aggregate scores. Due to this, many important attributes of the paper and pencil data sets were absent from the analysis process.

The back-to-back stem-and-leaf graph that one pair created in paper and pencil had its foundations in exploratory data analysis. The stem-and-leaf graph was developed by John Tukey (1977) who is credited for the development of exploratory data analysis. The stem-and-leaf display is considered valuable since it retains all of the observed values in rank order, but it also conveys the shape of the distribution, encouraging exploration of data. The stem-and-leaf display is a simplification of the data, but the numeric values retain their original meaning thus minimizing the chance of oversimplification. The stem-and-leaf display was the representation created in paper and pencil that was the most similar to the representations created in Tinkerplots. It was very similar in structure to the dot plots created in Tinkerplots.

**Method of analysis.** There was a difference in the pairs' method of analysis in the two environments. In paper and pencil, the pairs' quantified the data. The most

common representation in paper and pencil was fractions. The pairs lumped all the data into two fractions and then compared their size. This was a very practical way to deal with groups of unequal size. Combining the data for each group of students into an aggregate score made it possible to compare the two groups of students on an equal basis. Thus, the creation of fractions helped them overcome the difference in group sizes. After the fractions were created, the pairs only had a few numeric quantities to analyze.

In TinkerPlots the pairs never quantified the data like they did in paper and pencil. Rather, their analyses were based on a visual inspection of the graphs and qualitative comparisons. The subjects were able to visualize the entire range of the data arranged from least to greatest, and how the scores of both groups of students were placed throughout. The pairs were able to perform a side-by-side comparison of the two groups of students. These visualizations helped the pairs analyze the placement of the test scores within the distribution and in relation to each other.

Each of the pairs in TinkerPlots made observations regarding the placement of the two groups of students in relation to one another on the graph. These insights were a major focus in their analysis of the data. Three related statistical thinking elements demonstrated throughout the TinkerPlots sessions that relied on visual inspection were spread, modal clumps, and distribution chunks. All of the pairs in TinkerPlots recognized and analyzed distribution chunks within the data. The only two pairs in paper and pencil that used any of these statistical thinking elements were also the only two pairs that created graphical representations of the data.

In TinkerPlots, there was more of a focus on extreme scores. This was significant because extreme scores played a role in the hypotheses and final conclusions. This was

not the case in paper and pencil as extreme scores did not play a role in their analysis. The extreme low score (which was also an outlier) was frequently paired with conjecturing about causal relationships. In TinkerPlots, the pairs could see the outlier and felt a need to explain it. The outlier was easily recognizable and prominent due to the contrasting colors and sequencing of the data from least to greatest.

In paper and pencil, however, the outlier was not visible in their representations. Four of the pairs represented the data by combining the student scores into aggregate quantities. Another pair removed the extreme scores from the data set. Being absorbed into the data set, or deleted, the pairs did not have the opportunity to focus on the extremes. Thus, they were not included in the pair's hypothesis and conclusions.

***Relationship to related literature.*** Much of research in comparing data sets has focused on the average (Mokros & Russell, 1995; Watson & Moritz, 2000; Strauss & Bichler, 1988). In my study measures of center were used in only three sessions; two TinkerPlots sessions and one paper and pencil session. Hammerman and Rubin (2004) also found that using measures of center was not a very common in TinkerPlots when comparing groups.

An important step in considering data from a statistical perspective is to make a shift from thinking about the data in local ways to considering the data as a whole entity (Ben-Zvi, 2004, Konold & Higgins, 2003). In my study, in TinkerPlots, the participants were able to view the data both locally and globally. They viewed individual data points, focusing on the extreme, but also analyzed the entire range of the data. Within the graphs where the whole distribution was visible from least to greatest, they were also able to click on individual pieces of data and view their details on the data card. The significance

of this functionality is supported by Ben-Zvi and Arcavi (2001) who concluded that it was important that students be able to move back-and-forth between the two both views of the data.

Several studies have explored the types of strategies that students use when comparing data sets. Gal, Rothschild, and Wagner (1990) examined the techniques that third, sixth, and ninth grade students use to compare pairs of data sets of unequal size, with an emphasis on the average and how it was used. The algorithm for the average was used by 4% of the third graders, 14% of the sixth graders, and 48% of the ninth graders. The researchers found that the students who did not use the arithmetic mean did not use other techniques such as proportional reasoning or visual comparison to reach correct conclusions. They concluded that “that children who did not use the arithmetic mean for the most part had no other ways available to them to derive correct solutions” (p. 8). As compared to this, I found that students used a variety of strategies to compare data sets.

The findings of my study were consistent with the work of Watson and Moritz (1999) who interviewed 88 students in third through ninth grade who were given two data sets in graphical form to compare. The students used visual, numerical, or a combination of the two types of strategies to compare the data sets. In the paper and pencil environment of my study, the strategies that the students used were the same as those reported by Watson and Moritz. One pair of students used only visual methods, four pairs used only numerical methods, and one pair used a combination of visual and numerical methods.

Outliers are individual data points that deviate from the trend, falling outside the overall pattern of the graph. Outliers are inevitable and cannot be avoided, occurring in

almost all data collection. Experts flexibly deal with outliers in different ways (Moore & McCabe, 1993). For example, outliers may be evidence of an extraordinary event. Or in other cases they may be a mismeasurement and appropriately deleted from the data set.

In my study, five of the pairs working in paper and pencil combined the data into aggregate quantities. The outliers in these situations were absorbed into summary totals. The other pair deleted the outliers from the graph they created. Therefore the outlier was not visible to any of the pairs and not included in their hypothesis and conclusions.

Before removing outliers, researchers should try to figure out why they are there and investigate them carefully (Moore & McCabe, 1993) Outliers should always be identified and discussed – never ignored. It is important to investigate the nature of the outliers before deciding how to handle them and represent the data appropriately considering they are an important part of the data set. Outliers can significantly distort data. When not interpreted the meaning of the data can be significantly altered. This suggests that the students working in paper and pencil in my study missed a very important step in their interpretation of the data. By deleting the outliers and absorbing them into the data set, they may have overlooked information important to the problem under investigation.

**Difference in group size.** There were differences in how the pairs dealt with the difference in group size between the two environments. In paper and pencil, the difference in group size was dealt with by the type of representation whereas in TinkerPlots, it was dealt with using visual analysis techniques.

In paper and pencil, some of the pairs stated that their method of representation was used to account for the difference in group size. Most of the pairs recognized the



difference in group size and their method of analysis accounted for it. Five of the pairs lumped the data for each group of students together into an aggregate score making it possible to compare the two groups of students on an equal basis. Once their representation was created, the difference in group size never played a part in their analysis.

All of the pairs in TinkerPlots recognized the difference in group size, making statements regarding it. This awareness affected all of the pairs' analysis of the data. Each of the pairs could see that the number of students who listened to classical music was different than the students that did not listen to classical music. The difference in group sizes was easily recognizable in the visual provided by each of the graphs. In contrast, the difference in group size in the paper and pencil environment played a part in the type of representation they decided to create.

*Relationship to related literature.* Multiplicative reasoning has been established as an important factor in the comparison of data sets of unequal size (Cobb, 1999; Watson & Moritz, 1999). Research has found that students who are multiplicative reasoners are more successful at comparing data sets that are unequal size. Students who are not multiplicative reasoners struggle comparing groups of unequal size. These students are much more likely to employ additive reasoning as a strategy, comparing numbers of students as opposed to the frequency of students. My findings indicate that in TinkerPlots the participants were able to compare data sets of unequal size without the use of multiplicative reasoning using qualitative techniques and visual analysis.

### **Implications and Significance of the Study**

In this section, implications and significance of the study will be discussed in

regards technology and instruction. Also, the interplay between the theoretical perspectives and the findings of this study will be discussed. Specifically, whether TinkerPlots functioned as a Mindtool and became an intellectual partner with the participants.

### **Implications of the Technology**

Personal computers clearly have great potential as classroom tools (Moore, 1997). The integration of technological tools in the classroom can have a major impact on student learning. They can promote a “richer, powerful, and flexible learning environment in which students are active learners of statistics” (Ben-Zvi, 2000, p. 149).

The purpose of this study was to gain a better understanding of how students think in a technological environment. This was accomplished by exploring the differences in the thinking of students while they work in a technological environment and comparing this to their work in a paper and pencil environment. The student work in paper and pencil provided the foundation from which to explore student thinking while working in a technological environment. The intent was to *tease* out how students think while in a technological environment.

There were affordances of working in paper and pencil. Representing the data by creating fractions was very creative. Traditional mathematics and statistics textbooks do not introduce students to this method of analysis. It was also a very practical way to deal with groups of unequal size. Combining the data into aggregate scores made it possible to compare the two groups of students on an equal basis; there were no longer unequal group sizes to deal with. Thus, the creation of fractions helped them overcome the difference in group sizes. The pairs did not have to analyze or interpret extreme scores;

through their choice of representation, or modification of, the extremes had been absorbed or removed from the data set.

General findings across all pairs working in TinkerPlots illustrated the affordances of TinkerPlots. The pairs created many representations that were complete and representative of the data, modifying them with ease. All the pairs added context to the data, including both concretization and conjecturing about causal relationships. Every pair demonstrated an analysis cycle involving analyzing the representation they created, stating a hypothesis about which group of students did better, and attempting to verify the hypothesis. The pairs' analysis was based on visual inspection of the graphs and qualitative comparisons. Each of the pairs made observations regarding the placement of the two groups of students in relation to one another on the graph using modal clumps and distribution chunks. All pairs were aware of the difference in group sizes and used visual analysis techniques to deal with it.

Additional affordances of TinkerPlots were a prolonged analysis that resulted in a more refined hypothesis. In TinkerPlots the pairs' responses were rich and full of context, with caveats, explanations and rationalizations. As compared to paper and pencil, their conclusions were more nuanced, taking into account more variables. TinkerPlots appears to be a valuable asset in enhancing student's statistical thinking.

**Did TinkerPlots function as a Mindtool and become an intellectual partner with the students in this study?** During the course of this study, each student pair entered into a relationship with TinkerPlots. Each student pair understood that TinkerPlots had certain responsibilities while they had theirs. In these intellectual partnerships, TinkerPlots was responsible for performing statistical functions while the

student pair was responsible for interpreting, judging, and relating to the information presented to them by the software.

It was evident in this study that the partnerships formed between student pairs and TinkerPlots allowed students to more freely discuss, argue, and build consensus about the data presented to them by the software than what was observed in the paper and pencil sessions.

One distinction that describes the intellectual partnership that formed between Tinkerplots and the pairs of students comes from Thomas Reeves (1998) who compares learning “from” computers to learning “with” computers. When learning “from” a computer the computer acts as a tutor serving as an instructional delivery system to improve basic skills. When learning “with” a computer the computer serves as resources to help build creativity and higher order thinking skills. This resonates with Salomon’s (1990) description of technology acting as an intellectual partner. In my study, Tinkerplots did help build student’s critical and statistical thinking.

In Tinkerplots the software did the heavy lifting, in the paper and pencil sessions both students were responsible for performing the tasks that Tinkerplots performed. Using Tinkerplots the students were learning “with” the technology which allowed them to spend the majority of their time developing strategies for solving complex problems and deeper understandings of the subject matter. Mindtools such as TinkerPlots, allow students to form intellectual partnerships that allow them to access a richer view of the data.

“Technology is most powerful when used as a tool for problem solving, conceptual development, and critical thinking” (Ringstaff & Kelley, 2002, p. 5). Teachers

and students manage the teaching and learning, rather than the technology controlling it.

“The overriding message that can be gleaned from most current research on the implementation of computer-based technology in K-12 education is that technology is a means, not an end; it is a tool for achieving instructional goals, not a goal in itself” (p. 1).

For technology to improve education several conditions are necessary:

- Technology is best used as one component in a broad-based reform effort.
- Teachers must be adequately trained to use technology.
- Teachers may need to change their beliefs about teaching to support student collaboration, inquiry, and interactive learning.
- Technology resources must be sufficient and accessible.
- Effective technology use requires long-term planning and support.
- Technology should be integrated into the curricular and instructional framework. (pp. 2-3)

In classrooms where advanced uses of technology are prevalent, teachers report they are encouraged to explore different perspectives and are more willing to experiment (Trilling & Hood, 1999). In such classrooms, students actively participate, spending less time on drill and practice and curiosity is encouraged. Technology also helps fosters student collaboration when they participate in project-based learning where higher-order thinking skills are encouraged.

While my study supports the argument that computer tools can become an intellectual partner with the students, the unfortunate reality is that teachers are not using computers in the most effective ways. In Wenglinsky’s (2005/2006) examination of the National Assessment of Educational Progress (NAEP) from 1996, 1998, and 2001 showed that using computers for solving complex problems produced greater benefits than using them for drill-and-practice. However, it was reported that only 30% of mathematics teachers used computers for tasks that involved higher-order thinking skills.

In fact, there is some evidence that technology is only being used superficially to teach statistics (Burill, 1997).

Educators must find ways to harness the power of technology to teach mathematics. Technology can help students explore mathematical ideas, enhance understanding of a broad range of problems types, foster intuition, and aid in the understanding of concepts (Hollenbeck, Fey, & Wray, 2010). Technology based tools have the potential to transform the learning experience for students (Ben-Zvi, 2000; Jonassen, 2000; NCTM, 2000; Salomon, 1990). However, to achieve the benefits from technology, it must be implemented and used responsibly by teachers. According to Hollenbeck, Fey, and Wray, “Teachers need to carefully select and design learning opportunities for students where technology is an essential component in developing student’ understanding, not where it is simply an appealing alternative to traditional instructional routines” (p. 275). Educators must carefully design mathematics instruction where technology is an integral part of the learning process. To do this effectively, educators must have access to training and professional development where they themselves can learn how technology can be used most effectively and where they can gain a better understanding of the roles of teacher and student in technology-centric learning environments.

### **Implications for Instruction**

Makar and Confrey (Makar, 2004; Makar & Confrey, 2002a, 2002b, 2004, 2005) found that teachers struggle with comparisons between distributions even when they use formal measures. In paper and pencil, I observed that the pairs had a difficult time comparing the data sets. Two of the pairs in paper and pencil created representations with

errors in them, and one pair's representation was accurate, but not representative of the entire data set. Creating a representation in paper and pencil was a very time intensive process leaving little time for analysis. Several of the pairs continued to analyze the data beyond their initial hypothesis due to a lack of understanding of the effect of different group sizes.

It is important to build early experiences in the comparison of data sets. Research suggests that in about third grade, children can begin activities involving the comparison of two equal sized data sets (Watson & Moritz, 1999). At this grade level instruction and learning experiences based on visual comparison are recommended. Activities for comparing data sets should continue through middle and secondary school. It is possible for older students to develop complex statistical skills while working in the context of the comparison of groups. They should be given a variety tasks that allow them to work with data sets of both equal and unequal sizes. Both visual and numerical strategies should be discussed and encouraged in techniques such as the creation of graphs or using measures of central tendency to summarize the data. Increasing complexity of techniques for the comparison of data sets should be fostered by classroom discussion about tasks developed within the curriculum.

The pairs' ability to compare distributions with unequal sizes was enhanced by the use of TinkerPlots. In TinkerPlots all the pairs were able to compare data sets of unequal size using qualitative techniques and visual analysis.

Context can allow students to explore ideas in situations that are both rich and meaningful to them. Context played an important role in the solution of the tasks. In TinkerPlots, the use of context in the form of conjectures about causal relationships were

the participants way of justifying their conclusion that listening to classical music did have an effect on test scores, despite this group of students containing the outlier. This illustrates the importance of engaging students in tasks comparing data sets that are context-specific.

In TinkerPlots, extreme scores played a role in the pairs' analysis of the task. The outlier was highly visible and the pairs felt a need to explain it. This was accomplished through the use of context, specifically in the form of casual relationships. For the majority of pairs in paper and pencil, the outlier was not visible as they had combined the data into aggregate. In the representation where it was visible, the pair deleted it from the data set. This suggests that in the classroom, a discussion about the meaning and interpretation of extreme scores could be beneficial.

In Tinkerplots, the participants were able to compare data sets of unequal size without the use of multiplicative reasoning using qualitative techniques and visual analysis. This is significant since multiplicative reasoning has been established as an important factor in the comparison of data sets of unequal size (Cobb, 1999; Watson & Moritz, 1999). Research has found that students who reason multiplicatively are more successful at comparing data sets that are unequal size, while students who do not reason multiplicatively rely on additive reasoning strategies and have more difficulty with these comparisons. Being able to reason multiplicatively is a more advanced skill. Bypassing this type of reasoning can give students the ability to successfully compare and analyze data sets, who would not be able to otherwise. For instruction, this would mean that comparing data sets of unequal size could be included into the curriculum for a wider range of students with differing abilities.



While my study supports the argument that computer tools can become an intellectual partner with the students, the unfortunate reality is that teachers are not using computers in the most effective ways. In Wenglinsky's (2005/2006) examination of the National Assessment of Educational Progress (NAEP) from 1996, 1998, and 2001 showed that using computers for solving complex problems produced greater benefits than using them for drill-and-practice. However, it was reported that only 30% of mathematics teachers used computers for tasks that involved higher-order thinking skills. In fact, there is some evidence that technology is only being used superficially to teach statistics (Burill, 1997).

Educators must find ways to harness the power of technology to teach mathematics. Technology can help students explore mathematical ideas, enhance understanding of a broad range of problems types, foster intuition, and aid in the understanding of concepts (Hollenbeck, Fey, & Wray, 2010). Technology based tools have the potential to transform the learning experience for students (Ben-Zvi, 2000; Jonassen, 2000; NCTM, 2000; Salomon, 1990). However, to achieve the benefits from technology, it must be implemented and used responsibly by teachers. According to Hollenbeck, Fey, and Wray, "Teachers need to carefully select and design learning opportunities for students where technology is an essential component in developing student' understanding, not where it is simply an appealing alternative to traditional instructional routines" (p. 275). Educators must carefully design mathematics instruction where technology is an integral part of the learning process. To do this effectively, educators must have access to training and professional development where they themselves can learn how technology can be used most effectively and where they can

gain a better understanding of the roles of teacher and student in technology-centric learning environments.

### **Limitations and Constraints**

All research studies have limitations. While I do think that Jonassen's overarching theory as technology functioning as a Mindtool is viable and I observed participants in the study enter into an intellectual partnership with TinkerPlots, there were constraints using Jonassen's Critical Thinking Model (2000) to analyze the data.

For the analysis of research question one, I modified Jonassen's Critical Thinking Model (2000) to analyze and evaluate the data. A substantial number of modifications to this framework had to be made. Many of the thinking elements in this model were not observed in this study. Different factors played a part in the absence of these elements, such as the question not being conducive to eliciting these types of thinking processes. Other thinking elements were too pervasive and embedded in what the students were doing to be able to code them in a meaningful way. During analysis, the critical and creative thinking were merged into one category due to similarities that existed between the thinking elements.

However, there were some advantages to Jonassen's Critical Thinking Model (2000). It allowed me to observe participant uses of context knowledge subdivided into two different categories, concretizing and casual relationships, illustrating the difference between them. It also allowed me to view the structure of the analysis cycle using the thinking elements of hypothesizing and verifying.

Another limitation of this study was that the characterizations of student thinking were based on my observations of the participants' sessions solving the tasks in paper and

pencil and TinkerPlots. These characterizations could have been enhanced by interviewing the participants about their responses on the statistical tasks. Questions could have been asked that were designed to encourage students to elaborate on their thoughts and to reflect on the strategies of their partner. This may have helped shed light on the difference in thinking in the two environments.

### **Recommendations for Future Research**

Several researchers have identified that student use of context knowledge is as a valuable asset in statistical thinking and the solution of statistical tasks (Langrall, Nisbett, Mooney, & Janssem, 2011; Papanastasiou & Melietiou-Mavrotheris, 2008). Researchers may further want to address the role of context in the solution of statistical tasks in a technological environment. The categories of student use of context knowledge identified by Langrall, Nisbett, Mooney, and Janssem (2011) could provide a framework to evaluate student thinking while working in TinkerPlots.

Fitzallen (2013) identified three strategies that students use to access the functions within TinkerPlots to construct graphical representations. Future research may want to explore these strategies that students used in TinkerPlots to their work in paper and pencil. The strategies that Fitzallen identified could be a useful framework. In this study, it is not clear why three of the pairs in TinkerPlots decided to verify their hypothesis and continue analysis. Comparing student work from both environments in light of the strategies that Fitzallen identified may aid in understanding why students in TinkerPlots continue analysis.

### **Closing Thoughts**

The purpose of this study is to gain an in-depth understanding of how students

think in a technological environment. The intent is to characterize the types of higher order thinking and statistical thinking that students exhibit in technological and nontechnological environments. The relationships documented between working in a computer environment and paper and pencil environment will help to establish the benefit of using computers to teach and learn statistics.

The question of how valuable computer technologies are to the learning of statistics needs to be examined more closely. With a better understanding of the value of using computer technology teachers might be more inclined to use technology less superficially and in more meaningful ways. In addition, school systems will be able to make more informed decisions about technology purchases.

The findings of this study illustrate the affordances of working in both paper and pencil and TinkerPlots. Based on my findings TinkerPlots appears to be a valuable asset in enhancing student's statistical thinking.

## REFERENCES

- Achieve, Inc. (2004). *Ready or not: creating a high school diploma that counts*. American Diploma Project. Washington, DC: Achieve, Inc. Retrieved from Achieve, Inc. website: <http://www.achieve.org/files/ReadyorNot.pdf>
- Almeqdadi, F. (2000). The effect of using the geometer's sketchpad (GSP) on Jordanian students' understanding of geometrical concepts. *Proceedings of the International Conference on Technology in Mathematics Education (July 5-7, 2000) Beirut, Lebanon*.
- Ash, J. (2005). *The effects of computer-assisted instruction on middle school mathematics achievement*. Retrieved from <http://orchardsoftware.com/docs/OrchardstudyAsh.pdf>
- Bakker, A. (2002). Route-type and landscape-type software for learning statistical data analysis. In B. Phillips (Ed.), *Developing a Statistically Literate Society. Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS6, July, 2002) Cape Town, South Africa*. Voorburg, The Netherlands: International Statistical Institute. Retrieved May 5, 2013, from <http://www.fisme.science.uu.nl/publicaties/literatuur/4762.pdf>
- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64-83.
- Bakker, A., & Frederickson, A. (2005). Comparing distributions and growing samples by hand and with a computer tool. In W. J. Masalski (Ed.), *Technology-supported mathematics learning environments: Sixty-seventh yearbook of the National Council of Teachers of Mathematics* (pp.75-91). Reston, VA: National Council of Teachers of Mathematics.
- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147-168). Dordrecht, The Netherlands: Kluwer.
- Barak, M., Ben-Chaim, D., & Zoller, U. (2007). Purposely teaching for the promotion of higher-order thinking skills: A case of critical thinking – science. *Research in Science Education*, 37(4), 353-369.

- Barrow, L., Markman, L., & Rouse, C. E. (2008, August). *Technology's edge: The educational benefits of computer-assisted instruction*. (NBER Working Paper No. 14240). Retrieved from The National Bureau of Economic Research website: [http://www.nber.org/papers/w14240.pdf?new\\_window=1](http://www.nber.org/papers/w14240.pdf?new_window=1)
- Bayraktar, S. (2002). A meta-analysis of the effectiveness of computer-assisted instruction in science education. *Journal of Research on Technology in Education*, 34(2), 173-188.
- Becker, H. J., Ravitz, J. L., & Wong Y. (1999). *Teaching, learning, and computing: 1998 National Survey* (Report No. 3). Retrieved from University of California, The Center for Research on Information Technology and Organizations website: [http://www.crito.uci.edu/tlc/findings/special\\_report/participants\\_rev.htm](http://www.crito.uci.edu/tlc/findings/special_report/participants_rev.htm)
- Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical reasoning. *Mathematical Thinking and Learning*, 2(1/2), 127–155.
- Ben-Zvi, D. (2004). Reasoning about data analysis. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 121-146). Dordrecht: Kluwer.
- Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education: Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Bahia, Brazil*. Voorburg, The Netherlands: International Statistical Institute. Retrieved May 5, 2013 from [http://www.stat.auckland.ac.nz/~iase/publications/17/2D1\\_BENZ.pdf](http://www.stat.auckland.ac.nz/~iase/publications/17/2D1_BENZ.pdf)
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45, 35-65.
- Ben-Zvi, D., & Friedlander, A. (1997). Statistical thinking in a technological environment. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE round table conference* (pp. 45–55). Voorburg, The Netherlands: International Statistical Institute.
- Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3-16). Dordrecht, The Netherlands: Kluwer.
- Biehler, R. (1993). Software tools and mathematics education: The case of statistics. In C. Keitel & K. Ruthven (Eds.), *Learning from computers: Mathematics education and technology* (pp. 68–100). Berlin, Germany: Springer.

- Birks, M., & Mills, J. (2011). *Grounded theory: A practical guide*. London: SAGE Publications.
- Bright, G. W., & Friel, S. N. (1998). Helping students interpret data. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 63-88). Mahwah, NJ: Erlbaum.
- Burrill, A. (1997). Graphing calculators and their potential for teaching and learning statistics. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE Round Table Conference* (pp. 15–28). Voorburg, The Netherlands: International Statistical Institute.
- Camtasia (Version 1.0.1) [Computer software]. Okemos, MI: TechSmith.
- Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education, 1*(1), 1-12.
- Charmez, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage Publications Ltd.
- Chick, H. L., & Watson, J. M. (2001). Data representations and interpretation by primary schools students working in groups. *Mathematics Education Research Journal, 13*, 91-111.
- Churches, A. (2008). *Bloom's taxonomy blooms digitally*. Retrieved from <http://www.techlearning.com/article/8670>
- Clark, F. B., & Kamii, C. (1996). Identification of multiplicative thinking in children grades 1-5. *Journal for Research in Mathematics Education, 27*(1), 41-51.
- Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York, NY: Routledge.
- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning, 7*(1), 5-43.
- College Board (2006). *College Board Standards for College Success<sup>TM</sup>: Mathematics and Statistics*. Retrieved from College Board website: [http://www.collegeboard.com/prod\\_downloads/about/association/academic/mathematics-statistics\\_cbscs.pdf](http://www.collegeboard.com/prod_downloads/about/association/academic/mathematics-statistics_cbscs.pdf)
- Cousins, J. B., & Ross, J. A. (1993). Improving higher order thinking skills by teaching 'with' the computer: A comparative study. *Journal of Research for Educational Computing, 26*, 94–116.

- delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4(1), 55-82
- Fey, J. T., Hollenbeck, R. M., & Wray, J. A. (2010). Technology and the teaching of mathematics. In B. J. Reys & E. Reys (Eds.), *In Mathematics Curriculum: Issues, Trends, and Future Directions: Seventy-second Yearbook of the National Council of Teachers of Mathematics* ( pp. 265-275). Reston, VA.: NCTM.
- Einav, A., & Nenad, R. (2012). Teaching probability by using GeoGebra Dynamic Tool and implementing critical thinking skills. *Proceedings of the 4th World Conference on Educational Sciences (February, 2012) Barcelona, Spain*, 4943-4947. doi:10.1016/j.sbspro.2012.06.364
- Fickes, M. (2004). *How much does technology really cost? School Planning & Management*. Retrieved from [http://www.peterli.com/spm/resources/articles/archive.php?article\\_id=669](http://www.peterli.com/spm/resources/articles/archive.php?article_id=669)
- Fitzallen, N. (2007). Evaluating data analysis software. *Australian Primary Mathematics Classroom*, 12(1), 23-28.
- Fitzallen, N. (2013). Characterizing Students' Interaction with TinkerPlots. *Technology Innovations in Statistics Education*, 7(1), 1-18.
- Fitzallen, N., & Watson, J. (2010). Developing statistical reasoning facilitated by TinkerPlots. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. Retrieved May 5, 2013, from [http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8\\_8A4\\_FITZALLEN.pdf](http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_8A4_FITZALLEN.pdf)
- Fitzallen, N., & Watson, J. (2011). Graph creation and interpretation: Putting skills and context together. In J. Clark, B. Kissane, J. Musley, T. Spencer, & S. Thorton (Eds.), *Mathematics: Traditions and [new] practices. Proceedings of the Joint Conference of the Australian Association of Mathematics Teachers and Mathematics Education Research Group of Australasia*, (pp. 253-260). Sydney: MERGA.
- Friel, S. N. (2008). The research frontier: Where technology interacts with the teaching and learning of data analysis and statistics. In G. W. Blume & M. K. Heid (Eds.), *Research on technology and the teaching and learning of mathematics: Cases and perspectives* (Vol. 2, pp. 279-331). Greenwich, CT: Information Age Publishing, Inc.
- GAISE (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A curriculum framework for PreK-12 statistics education*. Retrieved from



The American Statistical Association website:  
[http://www.amstat.org/education/gaise/GAISEPreK12\\_Intro.pdf](http://www.amstat.org/education/gaise/GAISEPreK12_Intro.pdf)

- Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, 70(1), 1-51.
- Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 47-78). Dordrecht, The Netherlands: Kluwer.
- Gal, I., Rothschild, K., & Wagner, D. A. (1989). *Which group is better? The development of statistical reasoning in school children*. Paper presented at the meeting of the Society for Research in Child Development, Kansas City, MO.
- Gal, I., Rothschild, K., & Wagner, D. A. (1990). *Statistical concepts and statistical heuristics in school children: Convergence or divergence?* Paper Presented at the Annual Meeting of the American Educational Research Association, Boston, MA.
- Hammerman, J. K., & Rubin, A. (2003). *Reasoning in the presence of variability*. Draft paper presented at the Third International Research Forum on Statistical Reasoning, Thinking and Literacy, Lincoln, NE. Retrieved May 3, 2013, from [http://cehs.unl.edu/srtl/pdf/hammerman\\_rubin.pdf](http://cehs.unl.edu/srtl/pdf/hammerman_rubin.pdf)
- Hammerman, J., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal*, 3(2), 1-41. Retrieved May, 14, 2013, from [http://www.stat.auckland.ac.nz/~iase/serj/SERJ3\(2\)\\_Hammerman\\_Rubin.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Hammerman_Rubin.pdf)
- Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337-364.
- Harel, G., & Sowder, L. (2005). Advanced mathematical-thinking at any age: Its nature and its development. *Mathematical Thinking and Learning*, 7(1), 27-50.
- Hartwig, F., & Dearling, B. E. (1979). *Exploratory data analysis*. Newbury Park, CA: SAGE Publications Inc.
- Hawkins, A. (1997). Myth-Conceptions! In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE round table conference* (pp.1-14). Voorburg, The Netherlands: International Statistical Institute.
- Hopson, M. H., Simms, R. L., & Knezek, G. A. (2001/2002). Using a technology-enriched environment to improve higher-order thinking skills. *Journal of Research on Technology in Education*, 34(2), 109-119.

- Hsu, Y. (2003). *The effectiveness of computer-assisted instruction in statistics education: A meta-analysis* (Doctoral dissertation, University of Arizona). Available from Dissertation Abstracts International, 64, 1526.
- Institute of Education Sciences. (2009). *Intervention report*. Washington, DC: U.S. Department of Education.
- Johnson, Y., & Hofbauer, P. (2002). Describing middle school students' organization of statistical data. In D. S. Mewborn, P. Sztajn, D. Y. White, H. G. Wiegel, R. L. Bryant, & K. Nooney (Eds.), *Proceedings of the twenty-fourth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 1281–1291). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
- Jonassen, D. H. (2000). *Computers as Mindtools for schools: Engaging critical thinking*. Upper Saddle River, NJ: Merrill.
- Keller, S. (2006). *CPMP-Tools Software™* [Computer software]. East Lansing, MI: Michigan State University and Core-Plus Mathematics Project.
- Keltner, B. & Ross, R.L. (1995). *The cost of school-based educational technology programs*. Santa Monica, CA: Rand Corporation. Retrieved from [http://www.rand.org/pubs/monograph\\_reports/MR634/index2.html](http://www.rand.org/pubs/monograph_reports/MR634/index2.html)
- Key Curriculum Press. (2005). *Fathom™: Dynamic Statistics Software* (Version 2.0) [Computer software]. Emeryville, CA: Author.
- Knudsen, J., & Gallagher, L. (2010). Integration of technology, curriculum, and paraprofessional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal*, 47(4) 833-878. Thousand Oaks, CA: Sage Journals. Retrieved from <http://aer.sagepub.com/content/47/4/833>
- Konold, C. (2002a). Alternatives to scatterplots. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Association for Statistical Education.
- Konold, C. (2002b). No title. In M. J. Smith (Ed.), *Proceedings of the NSF K-12 Mathematics and Science Curricula and Implementation Centers*, Washington, DC, 68–73.
- Konold, C. (2002c). Teaching concepts rather than conventions. *New England Journal of Mathematics*, 34(2), 69-81.
- Konold, C. (2007). Designing a data analysis tool for learners. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 267-291). New York: Taylor and Francis.

- Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. D. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193-215). Reston, VA: National Council of Mathematics Teachers.
- Konold, C., & Miller, C. D. (2001). *TinkerPlots*. (Version 0.55). [Computer Software]. Amherst, MA: University of Massachusetts.
- Konold, C., & Miller, C. D. (2005). *TinkerPlots™: Dynamic Data Exploration* [Computer software]. Emeryville, CA: Key Curriculum Press.
- Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 55(4), 259-289.
- Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE round table conference* (pp. 151-167). Voorburg, The Netherlands: International Statistical Institute.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Developing a Statistically Literate Society. Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS6, July, 2002) Cape Town, South Africa*. Voorburg, The Netherlands: International Statistical Institute. Retrieved March 10, 2010, from <http://srri.umass.edu/files/konold-2002sum.pdf>
- Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment in education and training* (pp. 9-28). Hillsdale, NJ: Lawrence Erlbaum.
- Kulik, C. C., & Kulik, J. A. (1986). Effectiveness of computer-based education in colleges. *AEDS Journal*, 19(1-2), 81-108.
- Lamon, S. (1994). Ration and proportion: Cognitive foundations in unitizing and norming. In Harel, G., & Confrey, J. (Eds.), *The Development of Multiplicative Reasoning in the Learning of Mathematics* (pp. 89-120). Ithaca, NY: State University of New York Press.
- Langrall, C., Nisbett, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning*, 13(1-2), 47-67.
- Lemke, C., Quinn, B., Zucker, A., & Cahill, S. (1998). *An analysis of the status of the education technology availability and usage in the public schools of Virginia*. Santa Monica, CA: Milken Exchange on Education Technology. Retrieved from <http://eric.ed.gov/?id=ED430534/>

- Lesh, R. A., & Lehrer, R. (2000). Iterative refinement cycles for videotape analysis of conceptual change. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 665–702). Mahwah, NJ: Lawrence Erlbaum.
- Lowe, J. (2001/2002). Computer-based education: Is it a panacea? *Journal of Research on Technology in Education*, 34(2), 163-171.
- Lewins, A., & Silver, C. (2007). *Using software in qualitative research: A step-by-step guide*. London: SAGE Publications Ltd.
- Lou, Y., Abrami, P., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, 71(3), 449–521.
- Makar, K. (2004). *Developing statistical inquiry: Prospective secondary mathematics and science teachers' investigations of equity and fairness through analysis of accountability data*. Austin: University of Texas.
- Makar, K., & Confrey, J. (2002). *Comparing two distributions: Investigating secondary teachers' statistical thinking*. Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS6, July, 2002), Cape Town, South Africa. Voorburg, The Netherlands: International Statistical Institute.
- Makar, K., & Confrey, J. (2004a). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 353-374). Boston: Kluwer Academic.
- Makar K., & Confrey, J. (2004b). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal*, 4(1), 27-54. Retrieved from <http://www.stat.auckland.ac.nz/serj>
- Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal*, 4(1), 27-54. Retrieved from <http://www.stat.auckland.ac.nz/serj>
- Mann, D., Shakeshaft, C., Becker, J., & Kottkamp, R. (1999). *West Virginia's Basic Skills/Computer Education Program: An Analysis of Student Achievement*. Santa Monica, CA: Milken Family Foundation.
- McClain, (2002). *Uses of interactive minitools to explore authentic data with teachers to create individual and collective thinking*. In B. Phillips (Ed.), Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS6, July, 2002), Cape Town, South Africa. Voorburg, The Netherlands: International Statistical Institute.

- McClain, K., Cobb, P., & Gravemeijer, K. (2000). Supporting students' ways of reasoning about data. In M. J. Burke & R. F. Curcio (Eds.), *Learning mathematics for a new century: 2000 yearbook* (pp. 174–187). Reston VA: National Council of Teachers of Mathematics.
- McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics*, 45, 103–129.
- McGraw, T. M., Ross, J. D., Blair, B. C., Hambrick, K. A., & Bradley, K. L. (2000). *Educational software use: The 2000 report*. Charleston, West Virginia: AEL.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco: Jossey-Bass.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (3rd ed.). Thousand Oaks, CA: Sage.
- Meletiou-Mavrotheris, M., Papanastasiou, E., & Stylianou, D. (2009). Enhancing statistics instruction in elementary schools: Integrating technology in professional development. *Montana Mathematics Enthusiast*, 6(1/2), 57-78.
- Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26, 20-39.
- Mooney, E. S., Langrall, C. W., Hofbauer, P. S., & Johnson, Y. A. (2001). Refining a framework on middle school students' statistical thinking. In R. Speiser, C. A. Maher, & C. N. Walter (Eds.), *Proceedings of the twenty-third annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 437–447). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
- Moore, D. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.
- Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, 65, 123-165.
- Moore, D. S., Cobb, G. W., Garfield, J., & Meeker, W. Q. (1995). Statistics education fin de siècle, *The American Statistician*, 49(3), 250-260.
- Moore D. S., & McCabe G. P. (1993). *Introduction to the practice of statistics* (2nd ed.). New York: W.H. Freeman and Company.

- Moses, K. D. (2002, January). They cost more than you think! TechKnowLogia, 59-62. Retrieved from [http://www.techknowlogia.org/TKL\\_active\\_pages2/CurrentArticles/main.asp?FileType=HTML&ArticleID=369](http://www.techknowlogia.org/TKL_active_pages2/CurrentArticles/main.asp?FileType=HTML&ArticleID=369)
- National Assessment Governing Board. (2004). *Mathematics Framework for 2005 National Assessment of Educational Progress*. Retrieved from <http://www.amstat.org/education/gaise/4>
- 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 Education Technology Plan 2010. (2010). Retrieved from the U.S. Department of Education website: <http://www.ed.gov/technology/netp-2010>
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, DC: Authors.
- National Mathematics Advisory Panel (2008). *The final report of the national mathematics advisory panel*. Retrieved from the U.S. Department of Education website: <http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf>
- Paparistodemou, E. & Melietiou-Mavrotheris, M. (2008). Developing young students' informal inference skill in data analysis. *Statistics Education Research Journal*, 7(2), 83-106.
- Pea, R. D. (1985). Beyond amplification: Using the computer to reorganize mental functioning. *Educational Psychologist*, 20(4), 167-182.
- Pea, R. D. (1987). Cognitive technologies for mathematics education. In A. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 89-122). Hillsdale, NJ: Erlbaum.
- Pearson Digital Learning. (2006). *Educators use Successmaker digital courseware system to address individual needs*. Retrieved from <http://www.pearson.com>
- Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical thinking and learning*, 5(2/3), 131-156.

- Pfannkuch, M., & Rubick, A. (2002). An exploration of students' statistical thinking with given data. *Statistical Education Research Journal*, 1(2), 4-21.
- Polly, D. (2011). Technology to develop algebraic reasoning. *Teaching Children Mathematics*, (8), 472. doi:10.2307/41199752.
- Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal*, 5(2), 46-68.
- Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 201-226). Dordrecht, The Netherlands: Kluwer.
- Redhead, P. (2001). Investigating the total cost of technology in schools: Tools and strategies for managing technology investments. *Best Practices for Alberta School Jurisdictions*. Alberta Learning, Edmonton, Alberta. Retrieved from ERIC database. (ED467673)
- Reeves, T. C. (1999). A research agenda for interactive learning in the new millennium. In B. Collis & R. Oliver (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* (1999, pp. 15-20). Chesapeake, VA: AACE. Retrieved September 20, 2013 from <http://www.editlib.org/p/17393>.
- Ringstaff, K., & Kelley, L. (2002). *The learning return on our educational technology investment: A review of findings from research*. San Francisco, CA: WestEd RTEC.
- Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., ... Gallagher, L. P. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics three large scale studies. *American Educational Research Journal*, 47(4), 833-878.
- Rubin, A. (2002). Interactive visualization of statistical relationships: What do we gain? In B. Phillips (Ed.), *Developing a Statistically Literate Society. Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS6, July, 2002) Cape Town, South Africa*. Voorburg, The Netherlands: International Statistical Institute. Retrieved May 3, 2013, from [http://www.stat.auckland.ac.nz/~iase/publications/1/7f6\\_rubi.pdf](http://www.stat.auckland.ac.nz/~iase/publications/1/7f6_rubi.pdf)
- Rubin, A., & Hammerman, J. K. (2006). Understanding data through new software representations. In G. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth yearbook* (Vol. 68, pp. 241-256). Reston, VA: National Council of Teachers of Mathematics.

- Rubin, A., Mokros, J., & Friel S. (1996). Data: Kids, cats, and ads. A curriculum unit in *Investigations in number, data, and space*. Palo Alto, CA: Dale Seymour.
- Salomon, G. (1990). Cognitive effects with and of computer technology. *Communication Research*, 17(1), 26–45.
- Schacter, J. (1999). *The impact of education technology on student achievement: What the most current research has to say*. Santa Monica, CA: Milken Exchange on Education Technology.
- Science and Engineering Indicators: 2010. (2010). Retrieved from the National Science Foundation website: <http://www.nsf.gov/statistics/seind10/>
- Shaharuddin M. S., Zaidatun, T., & Nurbiha A., S. (2012). Web-based simulation learning framework to enhance students' critical thinking skills. *Procedia - Social And Behavioral Sciences*, 64(9), 372-381. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1877042812050215>
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester, Jr. (Ed.), *Handbook of research on mathematics teaching and learning* (2nd ed., pp. 957-1009). Greenwich, CT: Information Age.
- Shaughnessy, J. M., Ciancetta, M., & Canada, D. (2004, July). *Types of student reasoning on sampling tasks*. Paper presented at the 28th International Group for the Psychology of Mathematics Education, Bergen, Norway.
- Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (Part 1, pp. 205–237). Dordrecht, The Netherlands: Kluwer.
- Sivin-Kachala, J., Bialo, E. R., & Langford, J. (1998). *Report on the effectiveness of technology in schools, 1990-1997*. Washington, DC: Software Publisher's Association.
- Strauss, S. & Bichler, E. (1988). The development of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education*, 19, 64-80.
- Strauss, A. L. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA: Sage.
- Taylor, R. (1991). Teacher training curriculum project: Integration of computing into the teaching of secondary school science and art. *SIGCUE Outlook*, 21(1), 2–5.
- Taylor, C., & Gibbs, G. R. (2010). *How and what to code*. Retrieved from [onlineqda.hud.ac.uk/Intro\\_QDA/how\\_what\\_to\\_code.php](http://onlineqda.hud.ac.uk/Intro_QDA/how_what_to_code.php)

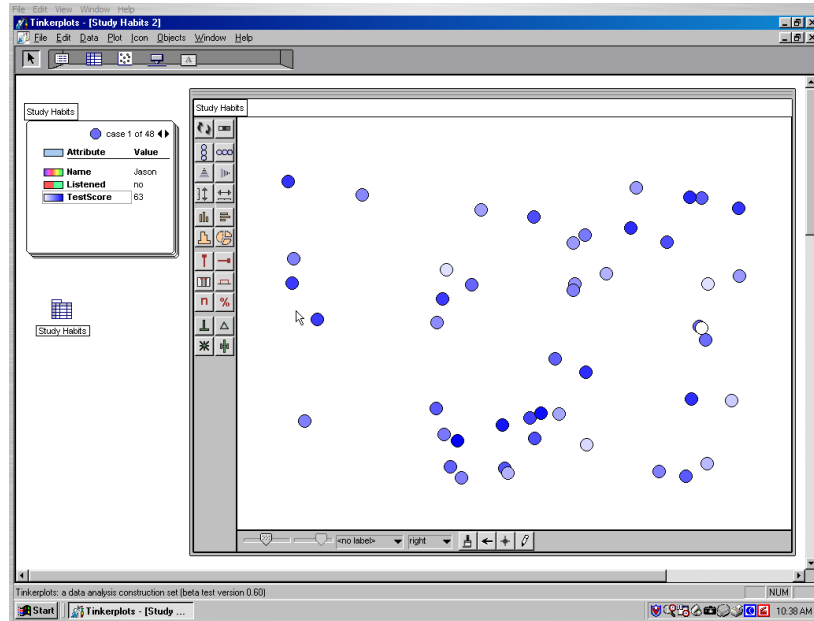


- Technology Counts '99. (1999, September 23). Building the digital curriculum. Report in collaboration with the Milken Exchange on Education Technology. *Education Week*, 59.
- Timmerman, C. E., & Kruepke, K. A. (2006). Computer-assisted instruction, media richness, and college student performance. *Communication Education*, 55(1), 73-104.
- Transana (Version 2.4) [Computer software]. Madison, WI: University of Wisconsin-Madison.
- Trilling, B., & Hood, P. (1999). Learning technology, and education reform in the knowledge age or “We’re wired, webbed, and windowed, now what?” *Educational Technology*, 39(3), 5-18.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley Publishing Company.
- U.S. Department of Education. (2008). The final report of the national mathematics advisory panel. Retrieved from <http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf>
- Watson, J. (2008). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal*, 7(2), 59-82.
- Watson, J. M., & Chick, H. L. (2001a). A matter of perspective: Views of collaborative work in data handling. In M. vanden Heuvel-Panhuizen (Ed.), *Proceedings of the 25<sup>th</sup> conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 407-414), Utrecht, The Netherlands: Freudenthal Institute.
- Watson, J. M., & Chick, H. L. (2001b). Does it help?: Collaboration during mathematical problem solving. *Hiroshima Journal of Mathematics Education*, 9, 33-73.
- Watson, J., & Donne, J. (2008). Building informal inference in grade 7. In M. Goos, R. Brown, & K. Makar (Eds.), *Navigating currents: Charting directions* (Proceedings of the 31<sup>st</sup> Annual Conference of the Mathematics Education Research Group in Australasia, Brisbane, pp. 555-562). Sydney, NSW: MERGA.
- Watson, J., & Donne, J. (2009). TinkerPlots as a research tool to explore student understanding. *Technology Innovations in Statistics Education*, 3(1)1-36. Retrieved May 3, 2013, from <http://escholarship.org/uc/item/8dp5t34t>
- Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, 34, 1-29.

- Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37, 145-168.
- Watson, J. M., & Moritz, J. B. (2000b). The development of concepts of average. *Focus on Learning Problems in Mathematics*, 11, 15-39.
- Watson, J., & Wright, S. (2008). Building information inference with TinkerPlots in a measurement context. *Australian Mathematics Teacher*, 64(4), 31-40.
- Wenglinsky, H. (1998). *Does it compute? The relationship between educational technology and student achievement in mathematics*. Princeton, NJ: Educational Testing Service Policy Information Center.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67, 223-265.
- Williams, G. D. (1996). *An examination of computer aided instruction, the self-concept and achievement levels of developmental mathematics community college students* (Doctoral dissertation). *Dissertation Abstracts International*, 57, 2402.
- Yin, R. K. (2011). *Qualitative research from start to finish*. New York, NY: The Guilford Press.

## APPENDIX A

### OPENING SCREEN OF TINKERPLOTS



## APPENDIX B

### STATISTICAL TASK: STUDY HABITS

Mrs. Jones was talking to the students in one of her mathematics classes about an article she read. The article described a group of children who performed better on tests because they had listened to classical music while studying. After hearing this, some of her students planned to listen to classical music while they studied for their next math exam.

To see how well students performed on this test, Mrs. Jones had her students mark an 'X' next to their name on the test if they listened to classical music while studying. The results of the 80-point test are listed below.

Analyze the data to see if students who listened to classical music while studying performed better on the math test than students who did not listen to classical music while studying.

1. How did the test scores for students who listened to classical music compare to the test scores of students who did not listen to classical music while studying?
2. What conclusion can you make about the effect of listening to classical music while studying?

Jason	63
Anna x	61
Lisa	48
Robert	57
Steven	70
Carla	72
Aaron x	68
Juan	73
Li	60
Terri	61
Casey	64
Shawn x	73
Kayla x	67
Portia	51
Derek	66

Mary	54
Dana	55
Laura	80
Mia x	77
Nina	60
David	63
Sean x	40
Alan	68
Cindy x	66
Dan	71
Melissa	59
Diane	52
Edward	45
Jamal x	71
Kathy	60

Barb	66
Marie	68
Phil x	78
Ian	56
Michael x	60
Linda	46
Cheryl	45
Stephanie x	65
Paul	56

Lee	74
Frances x	72
Latoya x	60
Gina	65
Ricky	52
Kaye	58
Ben	56
Chris	71
Toni	73

## APPENDIX C

### ISOMORPHIC TASK: DRIVERS' EDUCATION

Mr. Williams, a high school Drivers' Education instructor at Jefferson High School in Missouri, requested some information from the Missouri Department of Motor Vehicles. This information showed the scores of students from Jefferson High School on the 80-point written portion of the test who took the drivers exam last year. Some of these students took the Drivers' Education course and some did not because Missouri does not require all students to take Drivers Education courses.

Mr. Williams would like for you to analyze the data to see if students who took Drivers' Education scored higher on the drivers' examination than those who did not take Drivers' Education. The students who did take Drivers' Education have an "x" next to their name on the list below.

Analyze the data to see if students who took Driver's Education performed better on the test than students who did not take Driver's Education.

1. How did the test scores for students who took Driver's Education compare to the test scores of students who did not take Driver's Education?
2. What conclusion can you make about the effect of taking Driver's Education?

Jacob	53
Amanda x	51
Lori	38
Richard	47
Scott	60
Carrie	62
Adam x	58
Julio	63
Ling	50
Tammi	51
Caleb	54
Shane x	63
Katie x	57
Paige	41
Don	56
Becky	56
Marcie	58
Peter x	68

Megan	44
Deborah	45
Lindsey	70
Mya x	67
Nancy	50
Dale	53
Seth x	30
Alex	58
Carol x	56
Dennis	61
Monica	49
Dawn	42
Eric	35
Jovan x	61
Kelly	50
Luke	64
Frank x	62
Latisha x	50

Ira	46
Matthew x	50
Lynn	36
Connie	35
Sarah x	55
Patrick	46

Gail	55
Ryan	42
Kristin	48
Brett	46
Chad	61
Tori	63

## APPENDIX D

### DEFINITIONS OF CRITICAL AND CREATIVE THINKING:

JONASSEN (2000)

#### Critical Thinking

##### **Analyzing**

**Recognizing Patterns.** Students will perceive progressively more complex patterns of organization. Starting with simple repetition and recognition of visual or aural patterns, students should gradually master a variety of conventional or repetitive structures in many kinds of organized bodies of material, e.g. texts, works of art, quantitative data, or situations.

**Classifying.** Students will learn that elements can be sorted into groups with common attributes (categories). These categories can be labeled to reflect the common attribute. This process helps show how the parts can be grouped in meaningful ways which may help clarify organization.

**Identifying Assumptions.** Students will realize that many assertions, works of art, problem statements, contain hidden parts, underlying suppositions, or beliefs accepted to be true without the need of support. Since detecting such assumptions requires being able to recognize missing parts – things that are not stated but are implied by other parts of the whole—this is one of the more developmentally advanced skills. Teachers can help younger students by pointing out assumptions in statements made by the teachers and student.

**Identifying Main Ideas.** Students will be able to find the central meaning of a passage, a work of art, a set of qualitative data, or a situation. Doing this requires the ability to differentiate the core idea(s) from supporting or extraneous details, which is a hierarchical relationship. For this reason, identifying main ideas is relatively advanced, which may explain why many primary students find it difficult. Teachers can help young children by asking them what the main idea is of something they've said, and by asking them if something is a main (general) idea or an interesting (but minor) detail.

**Finding Sequence.** Students will determine the consecutive order of the parts of a sequentially organized body of material. At the simplest level they can order a groups of objects from smallest to largest or list the steps in which a simple process is performed.



At more sophisticated levels the necessary order of events can be determined even when not presented sequentially—a novel with flashbacks, for instance, or a math problem presented with extraneous information.

## **Connecting**

**Comparing/Contrasting.** Students will note similarities and differences between objects, events or actions. They many start with simple, global comparisons (That game is more fun than this one) but should eventually master complex, multi-attribute comparisons, such as comparing two theatrical productions or several business plans.

**Logical Thinking.** Students will learn to apply the rules of logic to analyze and accurately construct valid argument or conclusion. Logic is not a single skill but a complex system of skills. It is being treated as an entity here because there are many structured courses of study and materials available to teach it. Teachers can ask students to give reasons for statements they may make; they can also ask them to analyze other people’s arguments (assertions and support).

**Inferring Deductively.** Students will learn how to derive logical conclusions from accepted generalizations, truths, or principles. They will come to understand that if certain statements are accepted as true, certain others will necessarily also be true. This skill is a subcategory of logical thinking since it may involve syllogistic reasoning and rules of logical implication, but it seems worth treating separately because it is so important in informal reasoning and argument. If a certain rule is in effect and someone breaks it, then what must happen? Somewhat older students can understand the concept of things being true by definition.

**Inferring Inductively.** Students will learn how to develop a theory or draw a conclusion from empirical data. That is, after observing natural phenomena, surveying opinions, or studying statistical information, students will develop reasonable explanations or predictions. Unlike deductive inferences, which are logically necessary conclusions, inductive inferences yield probable conclusions. Because they are based on observation, it is always possible that new observations will make an inductive inference less convincing, as the history of science shows.

**Identifying Causal Relationships.** Students will be able to suggest causes of an event or situation (why it happened) and predict possible effects (what might happen as a result). Younger children may have difficulty with casual relationships but can be guided by questions such as “What happened right before this?” and “What usually happens when...?” Older students will need to distinguish between correlation (things that go together, perhaps by chance or because they are both caused by a third factor) and causality (things that make another thing happen). They will also need to understand multiple causes and chains of effects.

## **Evaluating**

**Assessing Information.** Students will appraise information and its sources as to reliability (to what extent one can believe it) and relevance (how it is connected to the purpose at hand). Students should learn that printed material is not necessarily accurate and learn some specific criteria to use in judging information, such as publication in a reputable periodical, qualifications and reputation of the author, evidence presented, and corroboration by other sources. They must also learn to decide if material fits their purposes—does it support their argument, is it relevant to the topic of a report, or is it interesting but not really connected.

**Determining Criteria.** Students will develop a basis for judging assertion, products, works of arts, etc. by establishing criteria or standards and a clearly defined process for applying them. Students should learn that many criteria (e.g., cost, effectiveness, appeal to certain groups, simplicity) can be used in evaluating and that the criteria selected will determine an evaluation’s outcome. They should also learn different ways of applying criteria, such as rank ordering, rating scales, quantitative measurement, etc.

**Prioritizing.** Students will order a set of options according to their importance. This can start with something as simple as making a class “to do” list each day and deciding which items are most important to get done. At a more sophisticated level, students must decide how they are defining “important” which involves determining criteria.

**Recognizing Fallacies.** Students will perceive errors in reasoning, including common logical fallacies such as vagueness, circular reasoning, nonsequiturs, and other propaganda techniques. This is a set of fairly advanced skills closely related to logical thinking in the analyzing sub-category. Students will have to be taught the various commonly recognized fallacies and why they represent incorrect reasoning. They will need considerable practice in identifying these patterns before recognizing fallacies becomes a part of their evaluating techniques. Many language arts, social studies, and math texts contain detailed descriptions of and exercises on the fallacies.

**Verifying.** Students will investigate the implication or results of a belief, position, hypothesis, or stance in order to confirm or reject it. Verifying involves what is sometimes referred to as “reality testing,” or evaluating an idea or plan against the criterion of feasibility or practicality. In some cases verifying can be done by careful cause/effect thinking; in other cases students can actually try something out and observe the effects.

## Creative Thinking

### **Elaborating**

**Expanding.** Students will add details, examples, qualifications, etc., to a core concept or principle. This skill builds on fluency and may be used in conjunction with a fluency

activity. Expanding ideas helps students understand and communicate them better. This skill is especially important in writing.

**Modifying.** Students refine or alter a core concepts, statement, principle, etc., in order to achieve different purposes. Students will need to be able to compare and contrast the requirements of different purposes so as to modify material appropriately. Revision in the writing process is an example of modifying.

**Extending.** Students will take principles, concepts, and conclusions from one context or frame of reference and apply them to another. This skill is similar to analogical thinking in that elements from different universes are combined. Making connections and seeing relationships between disparate areas often yield new, productive ideas.

**Shifting Categories.** Students will be able to change the direction of their thinking and take alternative points of view. This flexible thinking is very important in generating new ideas and in the complex thinking process. It is often linked with fluency in creativity theory and materials, and can be developed by asking students to think of items from many categories. It builds on the critical thinking skill of classifying, in that a familiarity with categories is implied.

**Concretizing.** Students will be able to make a general idea specific by giving examples and applications which will make an abstraction meaningful. This skill builds on the critical thinking skill of finding the main idea and requires understanding hierarchal relationships, so it is somewhat more developmentally advanced than the other elaborating skills. Like the other elaborating skills, it enhances both students' own understanding and their ability to communicate that understanding.

### **Synthesizing**

**Analogical Thinking.** Students will be able to use figurative language to express ideas in more vivid and novel ways. By using metaphors, similes, personification and other figures of speech, students can often make abstract or complex ideas clearer and gain insights into their meaning. Statements such as "Fear is a big black dog" followed by specific points of comparison can help students understand a concept in a concrete way and stimulate new ways of looking at the idea. Analogical thinking is different from comparing and contrasting in critical thinking because the things compared are of different types. Comparing involves two laws, or two types of literary genres, etc., whereas analogical thinking compares two seemingly very different types and finds surprising similarities (fear and dogs, life and a football game, etc.). Certain creative thinking programs such as Synectics are based heavily on analogical thinking.

**Summarizing.** Students will be able to produce a succinct form of a complex body of material. This requires them to identify the main idea and the basic structure and then express these elements more briefly than in the original. They must know what to leave out, which is often difficult for younger students. For many students, it will be necessary

for the teacher to model summarizing, both orally and by marking paragraphs to show important and nonessential information.

**Hypothesizing.** Students will learn to develop a testable explanation for a given situation or set of facts. This skill builds on inductive inferring, in students must make generalizations about observed data, but in hypothesizing the generalization (hypothesis) must be formulated so that it can be tested by experiment or structured observation (verifying). These skills used together are the basic skeleton of the scientific method, and most science textbooks contain teaching materials to develop them.

**Planning.** Students will be able to analyze a task and formulate a step-by-step procedure for accomplishing it. This requires deciding what the steps are and finding a reasonable sequence of performing the steps. Useful strategies are for teachers to model planning, and for students to share their plans. Planning may well be a skill related to personality style, with some people preferring detailed written plans and others operating well by planning in their heads.

### **Imagining**

**Fluency.** Students will generate many relevant responses when presented with a stimulus. The point of this skill is to produce many ideas so as to have a large body of material to work with. A basic principle underlying strategies to promote fluency is the deferral of judgment—accepts all responses at first, encourage many varied and unusual ideas. Evaluating is a different thinking process to be performed separately.

**Predicting.** Students will judge what things are likely to follow, be caused by, or result from a given situation or set of conditions. To do this, they must analyze the given situation and the future situation, compare these situations or conditions, and infer what may happen in the future because of what's happening now. Because the critical thinking skills implied in predicting, it is one of the more developmentally advanced creative thinking skills.

**Speculating.** Students will think and wonder about possibilities. Sometimes referred to as “what if” thinking, speculating may start from contrary-to-fact or whimsical conditions (what if everyone had the same name?) to stimulate novel ideas and insights that may then be applied to real situations. Unlike predicting, which seeks logically and factually probable ideas, speculating aims at generating interesting, unusual, perhaps humorous ideas. These may or may not lead to hypothesizing and verifying.

**Visualizing.** Students will think in mental images that they can later communicate. Despite the name, these images can reflect any of the sensory system—visual, aural, kinesthetic, tactile, olfactory. Visualizing can be used as a mental rehearsal (as in sports or performing arts), as a preparation for writing or an art activity, as an aid to observation and memory, and as a motivational tool (visualize what the completed project will look like and how it will feel to be done). A number on techniques and materials are available

to develop visualizing, sometimes found under the name of “guided fantasy” or “visual thinking.”

**Intuition.** Though not a skill that can be directly taught, intuition is a valuable mental process that students should recognize and have some notion of how to evoke. Intuition refers to flashes of insight, a seemingly instantaneous understanding without an awareness of sequential, rational thinking, a strong hunch about something without concrete evidence. Intuitions are often based on well-learned knowledge and skills which have sunk beneath the level of consciousness, and thus may be similar to what some refer to as automaticity. Students should see intuition as a type of personal knowledge that can serve as a starting point for more objective types of knowledge. Intuition often needs to be verified and subject to other types of critical thinking and may ultimately be rejected, but can also be the spark of truly original ideas.

## APPENDIX E

### “VARIATION-TALK”: ARTICULATING MEANING IN STATISTICS

MAKAR AND CONFREY (2004)

#### STANDARD STATISTICAL LANGUAGE

Proportion or Number Improved  
Mean  
Maximum/Minimum  
Sample Size  
Outliers and Extreme Values  
Range  
Shape (e.g., skewed, bell-shaped, normal)  
Standard Deviation

#### NONSTANDARD STATISTICAL LANGUAGE

##### Spread

Nonstandard language Makar and Confrey (2004) identified to describe spread: clustered, clumped, grouped, bunched, gathered, spread out, evenly distributed, scattered, and dispersed.

##### Low-middle-high

The distribution is subdivided into three parts: low, middle, and high (Bakker, 2004; Hammerman & Rubin, 2004).

##### Modal Clump

Konold et al., (2002) described ‘modal clumps’ as bumps were the middle hump of mound-shaped distributions. Students often used the mean as a way of characterizing a modal clump due to its location in the middle of the set. The modal clumps were generally ranges in the middle of the data sets that included the mode and had a higher percentage of data than either of the other two partitions. Konold et al., referred to the modal clump as a “range of values in the heart of a distribution” (p. 1).

## Distribution Chunk

In contrast to modal clumps, distribution chunks are groupings of data not in the center of the distribution. Recognition of distribution chunks involves discussion about groupings of the data including phrases such as “the bulk of them, the main group, the big chunk, and the majority” (Makar & Confrey, 2005).