5-2015

Differences in Statistical Reasoning Abilities through Behavioral-Cognitive Combinations of Videos and Formative Assessments in Undergraduate Statistics Courses

James M. Ramey
East Tennessee State University

Differences in Statistical Reasoning Abilities through Behavioral-Cognitive Combinations of Videos and Formative Assessments in Undergraduate Statistics Courses

---

A dissertation

presented to

the faculty of the Department of Educational Leadership and Policy Analysis

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Doctor of Education in Educational Leadership

---

by

James Michael Ramey

May 2015

Dr. Bethany Flora, Chair
Dr. Donald Good
Dr. Catherine Glascock
Dr. Karen King

Keywords: Multimedia learning, Formative assessment, Statistical thinking, Conceptual reasoning, Mastery learning

ABSTRACT

Differences in Statistical Reasoning Abilities through Behavioral-Cognitive Combinations of Videos and Formative Assessments in Undergraduate Statistics Courses

by

James Michael Ramey

This study evaluated whether significant differences in statistical reasoning abilities exist for completers of short online instructional videos and formative quizzes for students in undergraduate introductory statistics courses. Data for the study were gathered during the Fall 2013 semester at a community college in Northeast Tennessee.

Computer-based pedagogical tools can promote improved conceptual reasoning ability (Trumpower & Sarwar, 2010; Van der Merwe, 2012). Additionally, prior research demonstrated a significant relationship between formative quiz access and student achievement (Stull, Majerich, Bernacki, Varnum, & Ducette, 2011; Wilson, Boyd, Chen, & Jamal, 2011), as well as multimedia object access and student achievement (Bliwise, 2005; Miller, 2013). Four research questions were used to guide the study. A series of analysis of variance (ANOVA) statistical procedures was used to analyze the data.

Findings indicated no significant differences in statistical reasoning abilities between students who were provided access to supplemental online instructional videos and formative quizzes and

students who were not provided access. Moreover, statistical reasoning abilities did not differ significantly based upon number of quizzes successfully completed, average number of quiz attempts, or number of videos accessed.

DEDICATION

This study is dedicated to the memory of my father, James R. "Bud" Ramey. A born teacher and career educator, he appreciated learning for its own sake and always sought mastery in the many subjects that awakened his interest.

ACKNOWLEDGEMENTS

I first want to express my love and appreciation to my wonderful wife, Amy, and my dear girls, Angie Cate, Bess, and Joanna, for putting up with me during this arduous yet deeply rewarding process. Thanks also belong to my mother, Phyllis Murray Ramey, who is a continual source of support and encouragement.

I am also abundantly grateful to Dr. Bethany Flora for serving as my committee chair as well as to Dr. Donald Good, Dr. Catherine Glascock, and Dr. Karen King for agreeing to serve on my committee. Your questions, insights, and guidance have been invaluable. Thanks also go to Joanna Wicker and Betty Ann Proffitt who have always been willing to offer cheerful assistance.

Additionally, my abiding gratitude goes to Dawn Dabney for investing her time and expertise to assist me with the study and to Connie Church for providing helpful guidance and resources related to SPSS. Thanks also to Northeast State math faculty members Jan Lewenczuk, Cindy McNabb, and John Melendez for their participation. Finally, I extend my thanks to everyone who has supported me on the journey, especially good friends and colleagues like John Mark Grubb and Erika Adams, and my longtime supervisor and mentor, Dr. James C. Lefler.

TABLE OF CONTENTS

Page

ABSTRACT 2

DEDICATION 4

ACKNOWLEDGEMENTS 5

LIST OF TABLES 9

LIST OF FIGURES 10

Chapter

1. INTRODUCTION 11

Statement of the Problem 12

Purpose of the Study 12

Research Questions 14

Significance of the Study 14

Definitions of Terms 15

Limitations and Delimitations of the Study 17

Chapter Summary 18

2. REVIEW OF THE LITERATURE 19

Theoretical Background 20

Mastery Learning and the Feedback-Corrective Process 20

Formative Assessment 21

Multimedia Learning 23

Schema Theory 24

Conceptual Change 25
Statistical Reasoning 28
Computer-Assisted Instruction 31
Online Formative Assessment 33
Multimedia Interventions 34
3. METHODOLOGY 37
Research Questions and Null Hypotheses 38
Instrumentation 40
Sample 41
Data Collection 42
Data Analysis 43
4. FINDINGS 44
Demographics 44
Analyses of Research Questions 45
Research Question #1 45
Research Question #2 47
Research Question #3 51
Research Question #4 57
Chapter Summary 59
5. SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS 61
Summary of Findings 61
Research Question #1 62

Research Question #2 62

Research Question #3 63

Research Question #4 63

Recommendations for Policy ….. 64

Recommendations for Practice 64

Recommendations for Future Research 65

Conclusions 66

REFERENCES 67

APPENDICES 83

APPENDIX A: Permission to Conduct the Research 83

APPENDIX B: Demographic Survey 86

APPENDIX C: ARTIST Scale for Sampling Variability 87

VITA 95

LIST OF TABLES

Table Page

1. Sample Demographics 42

2. Means and Standard Deviations of the Control and Treatment Groups 46

3. Means and Standard Deviations of the Age Groups 48

4. Means and Standard Deviations of Males and Females 50

5. Means and Standard Deviations of Quizzes Completed 52

6. Means and Standard Deviations of Average Quiz Attempts 54

7. Means and Standard Deviations of Quizzes Completed and Average Quiz Attempts 56

8. Means and Standard Deviations of the Number of Video Tutorials Accessed 58

LIST OF FIGURES

Figure Page

1. Boxplot of ARTIST Scale Score Based on Group 47

2. Boxplot of ARTIST Scale Score Based on Age 49

3. Boxplot of ARTIST Scale Score Based on Gender 51

4. Boxplot of ARTIST Scale Score Based on Quizzes Completed 53

5. Boxplot of ARTIST Scale Score Based on Average Quiz Attempts 55

6. Estimated Marginal Means of ARTIST Scale Scores Based on Quizzes Completed and Average Quiz Attempts 57

7. Boxplot of ARTIST Scale Score Based on Number of Video Tutorials Accessed 59

CHAPTER 1

INTRODUCTION

Online tutorial sites and Massive Open Online Courses (MOOCs) appear frequently in the headlines (Christensen & Weise, 2014; Matthews, 2013; Swartz, 2015); as such the public's awareness of computer-assisted instruction (CAI) continues to grow. Particularly, the practice of combining short videos and formative quizzes for online instruction is increasingly more popular and expectations for academic outcomes have swelled (Kessler, 2011; Lewin, 2013; Noer, 2012). The combination of short online videos with formative quizzes serves as a supplement to enhance and support class activities.

The combination of videos and quizzes approximates an instructional feedback-corrective loop that correlates with significant gains in student achievement (Bloom, 1984; Kulik, Kulik, & Bangert-Drowns, 1990). Researchers often operationalize student achievement in terms of exam grades (Larwin & Larwin, 2011; Limniou & Smith, 2014; Sosa, Berger, Saw, & Mary, 2011). However, exam grades do not always capture student understanding of important concepts; students can use formulas to generate accurate results without understanding the meaning of the results and how they influence decision-making (Ben-Zvi & Garfield, 2004; Tugend, 2013).

Studies involving CAI have indicated that computer-based pedagogical tools promote improved conceptual reasoning ability (Trumpower & Sarwar, 2010; Van der Merwe, 2012). Implementing the feedback-corrective process via computer-based tools, then, offers the possibility to increase student achievement as well as promote conceptual reasoning skills. Conceptual reasoning abilities are a highly desired skill expressed by employers (NACE, 2013).

Statement of the Problem

Employment research indicates that recent graduates lack reasoning skills, particularly in the analysis of data and argument logic (Fischer, 2013; van Merriënboer & Kirschner, 2012). Moreover, sixty percent of employers indicate that recent graduates should possess problem-solving and decision-making skills, but only forty percent indicate that recent graduates adequately possess these skills (Maguire Associates, 2012).

The lack of reasoning abilities in the population of college graduates is evidenced within the field of statistics. Particularly, students successfully completing introductory statistics courses often lack corresponding conceptual reasoning abilities. Although conceptual reasoning is a primary learning outcome for statistics, students in introductory courses commonly persistent misconceptions (Ben-Zvi & Garfield, 2004; Castro Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007; Gal & Garfield, 1997; Zieffler et al., 2008). These misconceptions include errors about distributions (Bakker & Gravemeijer, 2004), variation (Reading & Shaughnessy, 2004), and sampling distributions (Chance, delMas, & Garfield, 2005). Such conceptual misunderstandings limit the ability to integrate principles and apply them to real-life scenarios (Ben Zvi & Garfield, 2004).

Purpose of the Study

Students in introductory statistics classes commonly hold misconceptions that persist (Ben-Zvi & Garfield, 2004; Castro Sotos et al., 2007; Zieffler et al., 2008). Examples of misconceptions include confusion about the validity of sample size and about the role of variance in sampling distributions (Kahneman, Slovic, & Tversky, 1982; Sedlmeier & Gigerenzer, 1997;

Tversky & Kahneman, 1971). Students demonstrate mastery of statistical calculations, but lack comprehension of underlying processes and therefore misinterpret the results (Chance et al., 2005).

These persistent misconceptions interfere with students' ability to reason properly about data and can therefore limit job performance following graduation (Fischer, 2013). While studies indicate the feedback-corrective process associated with mastery learning can promote academic performance (Bloom, 1984; Kulik et al., 1990), there have been no studies that examine the relationships between mastery learning and improved statistical reasoning ability.

Therefore, the purpose of this study was to evaluate whether significant differences in statistical reasoning abilities exist with the intervention of feedback and corrective mechanisms. Specifically, this study examined whether significant differences in statistical reasoning exist for completers of short online instructional videos and formative quizzes in undergraduate statistics courses. The dependent variable of statistical reasoning ability was operationalized as the score on the ARTIST Scale dealing with sampling variability. The Assessment Resource Tools for Improving Statistical Thinking (ARTIST) is a project funded by the National Science Foundation and supported by the Foundation along with the University of Minnesota and California Polytechnic State University (delMas, Garfield, Ooms, & Chance, 2007).

Research Questions

The following research questions were used to guide this study.

1. Is there a significant difference in mean ARTIST Scale score between students who were provided access to supplemental online instructional videos and formative quizzes and students who were not provided access?
2. Is there a significant difference in mean ARTIST Scale score as compared by student demographics?
   a. Is there a significant difference in mean ARTIST Scale score as compared by age?
   b. Is there a significant difference in mean ARTIST Scale score between males and females?
3. Is there a significant difference between mean ARTIST Scale scores when compared by number of formative quizzes successfully completed and average number of quiz attempts?
4. Is there a significant difference in mean ARTIST Scale score between students who opted not to access video tutorials and students who accessed the most video tutorials?

Significance of the Study

While numerous researchers have focused on the connections between multimedia learning or formative assessment and student achievement, few have examined the relationship between these pedagogical approaches and the conceptual abilities of students. As more employers begin to look beyond the grades of new graduates and set expectations about their abilities to interpret and reason about data, higher education institutions must adapt their modes of instruction to remove stubborn misconceptions. The misconceptions represent a failure of students to properly relate concepts to other concepts and topics within a broader field of knowledge (Chi, 2005; Özdemir & Clark, 2007; Posner, Strike, Hewson, & Gertzog, 1982).

Colleges and universities that adapt modes of instruction will equip students with the potential for greater success in their careers following graduation and, through course mastery, students can experience greater academic self-efficacy beliefs (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). These beliefs, in turn, correlate with college success (Gore, 2006). Furthermore, continued student success will contribute to retention and help institutions to remain competitive (Talbert, 2012).

Definitions of Terms

In order to clarify the meanings of terms used in the study, the following list offers selected definitions.

ARTIST: The Assessment Resource Tools for Improving Statistical Thinking (ARTIST) is a project funded by the National Science Foundation and features several scales that focus on specific statistical topics.

Conceptual Change: Conceptual change, as applied to education, is a fundamental shift in understanding wherein not only relationships between concepts are altered, but also relationships between individual concepts and the topic or subject as a whole (Chi & Hausmann, 2003; Posner et al., 1982; Vosniadou, 1994; diSessa, 1988).

Feedback-Corrective Process: The feedback-corrective process describes a technique in which learners are first confronted with timely information about their performance and then provided with additional instruction that addresses any gaps between their performance and that of mastery (Block, 1977; Block, 1980; Guskey, 1997).

Formative Assessment: Formative assessment is a means of evaluating student performance and providing feedback through quizzes and similar mechanisms that carry little or no weight relative to the final course grade (Guskey, 2010).

Misconception: A misconception represents an incomplete or otherwise incorrect understanding of a concept and how the concept relates to other concepts and topics within the broader academic subject (Chi, 2005; Özdemir & Clark, 2007; Posner et al., 1982).

Multimedia Learning: Multimedia learning is an approach in which instructional content is presented to learners though both the visual/pictorial and the verbal/auditory channels in order to maximize the limited capacity of the channels and the active processing abilities of the learners (Baddeley, 2007; Mayer, 2001; Paivio, 2006).

Sampling Error: Sampling error is the difference between the estimated value of a quantity in the sample and the true value of the quantity in the population (Witte & Witte, 2010).

Sampling Variability: Sampling variability describes the size of the sampling error across multiple hypothetical samples drawn from the same population (Witte & Witte, 2010).

Schema: A schema represents or models a learner's comprehension of a concept or group of concepts (Anderson, Spiro, & Anderson, 1978; Rumelhart & Norman, 1976).

Limitations and Delimitations of the Study

The current study was subject to both limitations and delimitations. Limitations are influences that place restrictions on the methodology and conclusions of a study and are beyond the researcher’s control, whereas delimitations are choices made by the researcher to set boundaries for the study (Creswell, 2013).

Limitations of the study arose from factors related to the sample. First, the study was limited to a single community college in the southeastern United States and involved seven sections of Probability & Statistics with four different faculty members. Therefore the results may not be generalizable to other community colleges in the United States, to other academic subject areas, or a single section of courses taught by only one faculty member.

Second, two faculty members instructed the treatment group and two other faculty members instructed the control group. Differing teaching styles and levels of experience could have influenced the results of the study. Third, data were collected for one semester only. Gathering data over two or more semesters would have generated a larger sample for the study.

Delimitations of the study included the literature selected for review, the population, and the methodological procedures chosen. First, the researcher omitted studies related to academic emotions and self-efficacy beliefs. While this research could have provided a basis for explaining the lack of participation in regard to the online instructional videos and formative quizzes, questions about participation levels were beyond the scope of the study.

Second, boundaries of the population were based upon a sample that had been gathered previously by community college instructors. Third, methodological procedures were likewise delimited based upon prior choices made by community college personnel. These choices include the research instrument, the online tutorial site, and the specific quizzes and videos made available.

Chapter Summary

The combination of short online videos and formative quizzes for instructional purposes has become more widely employed, in part, because some MOOCs and online tutorial sites have featured this pedagogical approach. The approach approximates the feedback-corrective process associated with mastery learning, and research has demonstrated that students instructed with the feedback-corrective process outperform students who are not.

In general, research related to mastery learning indicates that course grades do not always accurately gauge student conceptual mastery. Therefore, this study used the dependent variable of a nationally recognized statistical reasoning assessment as the measure of conceptual understanding. Moreover, short online videos and formative quizzes were the independent variables used as factors contributing to reduced misconceptions.

CHAPTER 2

REVIEW OF THE LITERATURE

With the increased media exposure allotted to Massive Open Online Courses (MOOCs) and tutorial sites like Khan Academy and Sophia.org, more attention has been given to the feedback-corrective process associated with mastery learning (Kessler, 2011; Lewin, 2013; Noer, 2012). Specifically, the feedback-corrective process in mastery learning has been compared to the combination of short online videos and formative quizzes central to the pedagogical approach employed in MOOCs and tutorial sites (Koller, 2012).

The purpose of this comparative study was to examine whether differences in statistical reasoning exist for students who use feedback and corrective mechanisms to resolve conceptual misunderstandings that are common to learners in statistics, especially misunderstandings about the validity of sample size and about the role of variance in sampling distributions (Kahneman et al., 1982; Sedlmeier & Gigerenzer, 1997; Tversky & Kahneman, 1971). In particular, this study examined whether significant differences in statistical reasoning exist due to the interventions of short online instructional videos and formative quizzes for students in undergraduate introductory statistics courses. For the purpose of this study statistical reasoning ability was defined as performance on a statistical reasoning assessment.

This literature review is presented in five sections. The first section establishes the theoretical background. The second section features a review of the literature regarding statistical reasoning. In the third section studies involving computer-assisted instruction (CAI) are reviewed. The fourth section presents a review of studies involving online formative assessment. The fifth and final section explores studies that involve multimedia interventions.

Theoretical Background

The theoretical background for this study was derived from five key areas: mastery learning, formative assessment, multimedia learning, schema theory, and conceptual change. Mastery learning, formative assessment, and multimedia learning theories supplied a basis for examining the multimedia intervention in this study, while schema theory and conceptual change yielded a framework for exploring how learners process and organize information.

Mastery Learning and the Feedback-Corrective Process

Mastery learning was first introduced by Bloom (1968) and later expanded as a strategy for employing feedback and corrective procedures to raise student achievement levels (Bloom, 1971). The performance gap that separated students could be narrowed "if every student had a very good tutor . . ." (p. 4). Bloom advocated a number of classroom techniques and procedures to approximate the benefits of one-on-one tutoring.

Bloom's research and thought regarding the benefits of mastery learning culminated in *The 2 Sigma Problem* (1984). In this seminal article, student learning was compared under three instructional conditions: conventional, mastery learning, and tutoring. The final achievement measures were striking. Using the standard deviation of the control group (the conventional class) as an index, the average test score for students in the tutoring group was two standard deviations above the control. Meanwhile, the average test score for students in the mastery learning group was one standard deviation above the control. While falling short of tutoring group students, students in the mastery learning group reached a level of achievement that far exceeded that of the control group. Later, a meta-analysis of 108 studies indicated that mastery learning was associated with a more modest improvement (0.5 standard deviations), but the researchers acknowledged that the effect was still "relatively strong" (Kulik et al., 1990, p. 292).

Students who engage in mastery learning receive feedback combined with corrective procedures that address misunderstandings and gaps in knowledge before a summative assessment is administered (Guskey, 2010). Following initial instruction, learners take a formative assessment that alerts them to areas for improvement. These areas for improvement are addressed with recommended activities, known as correctives, that present the material in a manner differing from that of the initial instruction (Block, 1977, 1980; Guskey, 1997). Following a period where learners apply correctives, a second test is administered. Students benefit from a second opportunity to succeed, and instructors benefit by gaining information about how helpful the correctives were (Guskey, 2007, 2010).

Elements of the feedback-corrective process draw from both formative assessment theory and multimedia learning theory. Formative assessment theory, in particular, supports a framework for considering elements that relate to feedback.

Formative Assessment

The term 'formative evaluation' originally referred to the evaluation of curricula (Scriven, 1967). Soon after, however, the idea of formative assessment was applied to student learning (Bloom, 1968). Instead of evaluating a program under development, the assessment would gauge a learner's grasp of concepts. By alerting students during the learning process, formative assessments help correct misunderstandings and fill in gaps in knowledge before students take a summative assessment. This was also a central element of Bloom's approach to mastery learning (Guskey, 2010).

Throughout the 70s, 80s, and 90s, formative assessment was explored in depth, culminating in a landmark meta-analysis of more than 250 studies (Black & Wiliam, 1998). An effect size of between .4 and .7 was shown to be achievable through formative assessment. More recently, a unifying framework was developed to integrate the various approaches (Black & Wiliam, 2009).

The main supporting elements of this framework were constructed from literature related to instructional feedback. There are three crucial processes that are closely connected to feedback. First, it must be established where learners are in their learning. Next, it must be determined where they need to be, and finally, it must be determined what needs to be done to get them to their destination. In every learning situation there is a gap which must be bridged, and feedback is the means of making the learner aware of this gap (Ramaprasad, 1983).

Drawing on a multitude of studies, Hattie and Timperley (2007) simplified feedback as: "information provided by an agent regarding aspects of one's performance or understanding" (p. 81). Additionally, "To be effective, feedback needs to be clear, purposeful, meaningful, and compatible with students' prior knowledge and to provide logical connections. It also needs to prompt active information processing on the part of learners, have low task complexity, relate to specific and clear goals, and provide little threat to the person at the self level" (p. 104).

Just as mastery learning feedback elements draw support from formative assessment theory, mastery learning corrective elements draw support from multimedia learning theory. Multimedia learning theory derives from research in both cognitive neuroscience and cognitive psychology.

Multimedia Learning

The Cognitive Theory of Multimedia Learning (Clark & Mayer, 2011; Mayer, 2005; Mayer, 2008) has three assumptions: 1) the dual-channels assumption; 2) the limited capacity assumption; and 3) the active processing assumption. The dual-channels assumption states that the human information processing systems consist of two channels: visual/pictorial and verbal/auditory, and the limited capacity assumption expresses that each of these channels has a limited capacity for processing. The active processing assumption declares that active learning involves coordinated processes which include the selecting of words and images, organizing words and images separately, and then integrating them together along with existing knowledge.

The dual-channels assumption is built on the prior work of Paivio (1971; 1986) and Baddeley (1986). Though these landmark studies were conducted at least three decades ago, both have been refined and updated more recently (Baddeley, 2007, 2012; Paivio, 2006). Support for the limited capacity assumption comes from cognitive load theory (Chandler & Sweller, 1991), wherein the limits of working memory are conceptualized in terms of cognitive load, implying that there is a maximum amount of information that a learner's short-term, working memory can handle at any one time. The active processing assumption involves three elements which include the selection of relevant information, the organization of selected information, and finally the integration of organized information (Mayer, 2001, 2005; Wittrock, 1992). Together, the dual-channels assumption, the limited capacity assumption, and the active processing assumption support the multimedia principle: “People learn more deeply from words and pictures than from words alone” (Mayer, 2001, p. 47).

Multimedia learning theory therefore supports the use of videos as mastery learning corrective elements and supplies a context for exploring how learners integrate new information. Schema theory provides additional support for understanding the integration process.

Schema Theory

“Schemas provide the basic unit of knowledge and through their operation can explain a substantial proportion of our learning-mediated intellectual performance” (Sweller, 1994). While the concept of a schema originated with Kant (1884), the term was introduced to the field of psychology in the context of how children make sense of new information (Piaget, 1926). Later, schemas were described as structures that exist in the memory (Bartlett, 1932). Bartlett viewed these structures as dynamic configurations composed of elements which would shift as circumstances dictated and which are useful for organizing past reactions in terms of given sets of conditions.

Configuration of these dynamic structures became associated with the three processes of schema development (Rumelhart & Norman, 1976). The first process is accretion or the simple accumulation of facts. When new pieces of information are encountered, they are incorporated into existing schemas. If no schemas can accommodate the new information, however, either the second process involving minor changes (tuning) or the third process involving major changes (restructuring) must take place. Tuning consists of replacing variables and/or constants in an existing schema, while leaving the basic relational structure intact. Restructuring, on the other hand, involves the generation of new schemas.

Anderson et al. (1978) further described a schema as “an abstract description of a thing or event. It characterizes the typical relations among its components and contains a slot or place holder for each component that can be instantiated with particular cases” (p. 314). Some key conclusions drawn from this study include the following: 1) interpreting a situation in terms of a schema means matching elements in the situation with the corresponding place holders in the schematic knowledge structure; 2) schemas that a person already possesses are a “principal determiner” (p. 434) of what will be learned; and 3) information that fits the schema is more likely to be learned.

Equipped with these particular dynamic features, schemas can serve as a context for investigating models of conceptual change. The literature related to conceptual change theory explains how schema elements relate to one another to produce an understanding (correct or incorrect) of a particular concept or group of concepts. Moreover, the literature reveals the nature of persistent misconceptions and how these misconceptions can be altered.

Conceptual Change

The origins of conceptual change theory are usually attributed to Kuhn (1970) and the phenomenon of change in scientific paradigms. These paradigms are frameworks of beliefs and assumptions about what the physical world is like. Kuhn noted that changing a paradigm required extensive, time-consuming reformulation of prior assumptions and re-evaluation of facts, and that furthermore, the established scientific community is resistant to this change. Because the work itself is challenging and the community resists any alteration of the comfortable status quo, shifting paradigms is a revolutionary phenomenon.

Whereas Kuhn looked at sweeping historical changes, Piaget focused on changes within an individual's personal conceptual framework. Accordingly, the theory of accommodation (Piaget & García, 1974) can be considered as a leading influence on conceptual change research in the field of cognitive developmental psychology and related fields. Over the next seven years researchers in science education began to note that students brought their own intuitive concepts into the classroom and that these concepts inhibited a correct understanding of things like force and energy (Driver & Easley, 1978; Novick & Nussbaum, 1981). Students would assimilate content received through instruction and combine with their native, intuitive concepts to produce what the literature refers to as 'misconceptions' (Brown, 1992; Castro Sotos et al., 2007; Chi, Roscoe, Slotta, Roy, & Chase, 2012; Vosniadou & Verschaffel, 2004).

Posner et al. (1982) argued that misconceptions were alternate frameworks and that in order to change them a learner must: 1) become dissatisfied with the old framework; 2) understand the new framework; 3) see the new as plausible; and 4) see the promise of the new in terms of how it can be applied.

In this approach to conceptual change an entire framework of ideas is shifted as a whole (Carey, 1985, 1999; C. Smith, 2007), and the shift can best be described as "theory change" (Carey, 1999, p. 292). According to a second approach known as "ontological shift" (Chi & Hausmann, 2003, p. 432), however, misconceptions result from learners classifying concepts in the wrong ontological category. A key example involves learners confusing an emergent process with a direct process (Chi, 2005). In this case learners may understand the direct process of blood circulation and then try to apply this to an emergent process like diffusion. Because the two processes occupy two different ontological categories, however, the challenge is to help learners construct a new category, and if necessary assign each process to the ontological place where it belongs (Chi et al., 2012; Chi & Roscoe, 2002; Chi, 2008; Chi, Slotta, & De Leeuw, 1994; Slotta, Chi, & Joram, 1995).

A third approach to conceptual change is composed of elements from the previously discussed approaches, but more emphasis is placed on the dynamic nature of conceptual models. Framework theory holds that the learner's ontological commitments can interfere with the processing of new information, such that a concept can be distorted to fit the pre-existing framework (Vosniadou & Verschaffel, 2004; Vosniadou, 1994, 2002; Vosniadou, Vamvakoussi, & Skopeliti, 2008). Conceptual change therefore involves altering the ontological commitments in order to correct any misconceptions.

The fourth major view considered is known as a "knowledge-in-pieces" (diSessa, 1988, p. 49) approach. In contrast with the other three views, where knowledge resides in a unified framework, the knowledge-in-pieces approach is one where knowledge resides in an "ecology of quasi-independent elements" (Özdemir & Clark, 2007, p. 351). Instead of being limited by a framework, knowledge is constrained by the complexity of a system composed of elements that interact in a dynamic, emergent manner (diSessa, 1993, 2002; diSessa, Gillespie, & Esterly, 2004). Misconceptions are therefore robust and stubborn because they are connected as part of a web of relationships. Accordingly, conceptual change occurs through reorganizing these relationships (Mayer, 2002).

The knowledge-in-pieces view, then, aligns more closely with the flexible, network structure associated with schema theory (Marshall, 1995) than any of the other three views. If learners make different connections between elements, schema can be altered. If learners make different connections between pieces of knowledge, misconceptions can be corrected and understanding can deepen. Because schemas are principle determiners of what students learn and are sensitive to changing the relationships between individual elements (Anderson et al., 1978; Marshall, 1995), a knowledge-in-pieces approach to conceptual change explains how changing the relationships between these elements can alter misconceptions and correct faulty reasoning.

Identifying what constitutes a misconception requires a clear delineation of concepts that form a correct understanding of a given topic. Accordingly, the current study depends upon a review of elements that contribute to a sound understanding of statistical reasoning processes in general and sampling distributions in particular.

Statistical Reasoning

Statistical reasoning involves making sense of statistical information and making interpretations from the data (Garfield, 2003). Unfortunately, misconceptions persist among students in introductory statistics courses (Ben-Zvi & Garfield, 2004; Castro Sotos et al., 2007; Zieffler et al., 2008), and especially in regard to sampling distributions (Chance et al., 2005; Saldanha & Thompson, 2002; Sedlmeier & Gigerenzer, 1997). Although students are able to make correct calculations, often they are unable to interpret the meaning of these calculations or comprehend underlying processes. “This stems from the notoriously difficult, abstract topic of sampling distributions, that requires students to combine earlier course topics such as sample, population, distribution, variability, and sampling" (Chance et al., 2005, p. 295).

In other words, students lack the ability to perform more advanced reasoning tasks because they have failed to successfully integrate these earlier topics. The integration of new information with prior knowledge is a crucial aspect of active processing (Wittrock, 1992) and schema creation or modification (Rumelhart & Norman, 1976). Therefore, it is important to consider the misconceptions that involve earlier, foundational ideas in order to understand why learners commonly experience difficulty with the complex concept of sampling distributions (Chance et al., 2005).

Three foundational ideas which learners need to grasp first in order to develop an adequate understanding of sampling distributions are sampling processes, characteristics of the different distributions, and the central limit theorem (Castro Sotos et al., 2007). Understanding sampling processes involves comprehending a balance between sample representativeness and variability. Sample representativeness refers to a similarity of characteristics between the sample and the population, whereas variability means that not all samples will bear a resemblance to the population in the same way or to the same degree. The representativeness heuristic, however, leads learners to believe that any given sample is representative of the population, regardless of the population's size. This misconception is called "belief in the law of small numbers" (Kahneman et al., 1982; Tversky & Kahneman, 1971).

Another misconception related to sample occurs when students apply the intuitive law of large numbers to sampling distributions the same as for frequency distributions. While the law of large numbers can be applied directly to frequency distributions, sampling distributions must have variance taken into account (Sedlmeier & Gigerenzer, 1997). Both misconceptions are connected to a fundamental misunderstanding of variability in random events. This idea is foundational to an adequate understanding of sampling processes, which is in turn foundational to an adequate understanding of sampling distributions (Castro Sotos et al., 2007).

If sample representativeness is misunderstood, such that learners ascribe the same shape and properties to both the sampling distribution and the population distribution, they will therefore confuse population distributions and sampling distributions. Further, by confusing the two distributions, while also misunderstanding the concept of representativeness, learners may not be able to distinguish between the distribution of a sample and the sampling distribution of a statistic (Chance et al., 2005).

Similar consequences result when learners hold misconceptions related to the central limit theorem. The central limit theorem states that, given a large enough sample size, the sampling distribution of the sample mean will approximate a normal distribution. One common misconception occurs when students interpret the central limit theorem to mean that the distribution of *any* statistic will approach a normal distribution as sample size increases while a related misconception occurs when students ascribe equivalence to an actual sampling distribution and the theoretical model of a normally distributed population (Batanero, Tauber, & Sanchez, 2004).

Because sampling processes, characteristics of different distributions, and the central limit theorem are essential to a sound understanding of sampling distributions, most of the barriers to understanding involve these topics. A key factor that hinders student comprehension involves confusion about the variability of random events (Castro Sotos et al., 2007; Chance et al., 2005). This factor affects student understanding of sample representativeness and by extension inhibits students from properly distinguishing between kinds of distributions and from comprehending the relevance of sample size for these distributions.

Studies involving computer-assisted instruction (CAI) have examined computer-based or Web-based technologies as a means of correcting these misconceptions and to enhance student academic performance. Meanwhile studies involving subcategories of CAI, specifically online formative assessment and multimedia interventions, have likewise examined the relationship between these technologies and student academic performance. Together, these empirical studies form a basis for comparing the feedback-corrective mechanisms employed in the present study.

Computer-Assisted Instruction

One recent meta-analysis involving computer-assisted instruction (CAI) demonstrated an effect size of .33, which indicated that a moderately strong relationship exists between CAI and performance on exams in statistics (Sosa et al., 2011). Studies involving the following characteristics were reviewed:

> Studies were included only if they met all of the following criteria: (a) they assessed the effectiveness of computer-based instructional tools in statistics, (b) they evaluated effectiveness based on objective performance or learning measures (i.e., test scores), (c) they provided comparison data from a lecture-based control group receiving instruction in statistics (thus, simple pre–post studies, studies in which all groups were exposed to computer-based tools, and studies in which the control group did not receive instruction were excluded), (d) the control

> group did not receive any form of computer-assisted instruction, (e) they provided sufficient information to calculate an effect size, and (f) they provided enough information for the authors to be able to rate the studies with regard to specific learner-centered features (p. 102).

In addition, students who received more instructional time via CAI outperformed students who received lecture-only instruction, and the large effect size (.97) indicates that the difference in performance can be accounted for by the additional instructional time.

A more extensive meta-analysis, involving 70 studies conducted between 1960 and 2010, revealed an effect size of 0.57 for the relationship between computer-assisted instruction and exam performance in postsecondary statistics (Larwin & Larwin, 2011). The results suggested that the typical student moved from the 50th to the 73rd percentile when computer-assisted instruction was deployed as part of the curriculum.

Another finding of the meta-analysis was that although CAI had no significant impact on student achievement until the 1980s, from the 1980s until the 2000s effect size increased. This corresponds with an increase in the availability and sophistication of computer hardware and software during the period (Cobb, 2007).

Along with superior test scores, computer-assisted instruction is associated with improved comprehension of conceptual relationships (Trumpower & Sarwar, 2010; Van der Merwe, 2012). In the Van der Merwe study, for example, the treatment group significantly outperformed the control group on a test about descriptive statistics and also on a later test about inferential statistics. Moreover, the treatment group significantly outperformed the control group on the Comprehensive Assessment of Outcomes in Statistics (CAOS) statistical reasoning assessment.

Computer-assisted instruction therefore has an empirically supported relationship with higher exam scores and superior conceptual reasoning ability. Two specific subsets of CAI that significantly increased student achievement include online formative assessments and multimedia interventions.

Online Formative Assessment

Empirical studies involving online formative assessments are commonly associated with increased student achievement, defined as exam scores and final course grades (Davis, 2013; Lawton et al., 2012; Limniou & Smith, 2014; G. Smith, 2007). Particularly, a significant relationship exists between the total number of quizzes accessed and student achievement (Stull, Majerich, Bernacki, Varnum, & Ducette, 2011; Wilson, Boyd, Chen, & Jamal, 2011).

Stull et al. (2011), for example, featured online formative quizzes that were made available prior to lecture. Performance on the posttest was positively correlated with the number of quizzes accessed. Along similar lines, Wilson et al. (2011) detailed the use of online software that presented formative quizzes. Significantly, students that completed more than twelve quizzes scored an average of ten percentage points higher on the midterm exam as opposed to students who completed twelve quizzes or less.

Therefore, a significant relationship exists not only between online formative assessments and student achievement in a broad sense, but also between student achievement and the number of assessments accessed. Studies involving multimedia interventions have indicated similar potential benefits in respect to student achievement.

Multimedia Interventions

Multimedia interventions are defined as Web-based or computer-based resources that deliver instructional content through audio and video channels. Studies involving multimedia interventions can be classified as: 1) a multimedia intervention is the primary means of instruction, or 2) the multimedia intervention is offered as supplemental instruction. Where the intervention is the primary means of instruction, a further distinction can be drawn between interventions that employ static visualizations and interventions that employ dynamic visualizations.

Static multimedia content is comparable to lectures in terms of test scores when the multimedia content is delivered as the primary form of instruction (Aberson, 2000; 2003). In contrast, dynamic multimedia visualizations are superior to traditional lectures in terms of exam scores when offered as primary instruction (Lloyd & Robertson, 2012; Shi, 2012). Similarly, multimedia interventions are associated with increased student achievement when offered as supplemental instruction, and this achievement has been indicated through exam scores as well as final course grades (Bastürk, 2005; Miller, 2013). Particularly, a significant relationship exists between the total number of multimedia objects accessed and student achievement (Bliwise, 2005; Miller, 2013).

Delivered as the primary means of instruction, multimedia interventions featuring static visualizations are comparable to lectures, with no significant difference between test scores of students viewing tutorials compared to students receiving content via traditional lecture format (Aberson et al., 2000; 2003). In contrast, multimedia interventions that feature dynamic visualizations are superior to lectures when delivered as the primary means of instruction (Gambari, Falode, & Adegbenro, 2014; Lloyd & Robertson, 2012; Shi, 2012). Two common forms of dynamic visualizations include screencast tutorials, which involves synching a voice recording with actions on a computer screen (Udell, 2004), and computer animations.

For example, students in an upper level psychology course in statistics who received their instruction via a screencast tutorial significantly outperformed the control group in terms of test scores (Lloyd & Robertson, 2012). Test scores were also significantly higher for students who received geometry instruction via computer animation compared to students who received traditional lecture-based instruction (Gambari et al., 2014). Similarly, beginning students in Chinese language who received instruction through computer animation demonstrated significantly greater mastery of Chinese characters (through means of posttest scores) compared to students who received traditional instruction (Shi, 2012).

Likewise, traditional lectures that are combined with a multimedia intervention are superior to lecture-only instruction (Bliwise, 2005). Students who took a traditional lecture course in statistics that was supplemented with multimedia tutorials significantly outperformed lecture-only students in terms of exam scores. Each tutorial was structured to subdivide content into smaller components and also to provide graphic representations as aids to comprehension. At the end of each tutorial, students were presented with a short quiz that assessed their mastery of concepts contained in the tutorial. The number of tutorial accesses was positively correlated with quiz performance and quiz performance, in turn, positively correlated with exam items.

Similarly, students who accessed at least 75% of available supplemental multimedia learning objects in an online course significantly outperformed students who accessed less than 50% of the learning objects in terms of final course grade (Miller, 2013). Twenty-two multimedia objects were included in the course and each related directly to course topics. Additionally, students were allowed to access all the objects throughout the duration of the course.

A significant relationship exists between multimedia interventions and student achievement regarding complex topics in various subject areas, including statistics. Likewise, a significant relationship exists between formative assessment and student achievement. Moreover, a significant relationship exists between use of supplemental computer-based instruction and the statistical reasoning ability of students. Nevertheless, no significant relationship has been demonstrated to exist between supplemental multimedia interventions and the statistical reasoning ability of students. The current study explores this relationship in the context of the feedback-corrective process from mastery learning, as approximated by the use of formative quizzes and short online videos.

CHAPTER 3

METHODOLOGY

The purpose of this study was to determine whether significant differences in statistical reasoning abilities exist for completers of short online instructional videos and formative quizzes in undergraduate statistics courses. This chapter describes the design of the proposed study, the research population, data collection and analysis procedures, and the research questions and hypotheses.

This study used a quasi-experimental research design; the control group and treatment group had approximately the same number of participants. A nonequivalent control group design was used because the proposed course sections constitute intact groups (Trochim & Donnelly, 2006). Educational researchers have noted the benefits of a quasi-experimental approach (Borman, 2002; Shadish, Cook, & Campbell, 2002; Slavin, 2008).

Three sections of an introductory statistics course at a community college comprised the control group while the treatment group consisted of four sections. All seven sections were taught by full-time faculty members. The control group sections and the treatment sections were assigned to meet for 80-minute sessions twice per week, either using a Monday and Wednesday or Tuesday and Thursday schedule. Additionally, the control and treatment sections participated in the same number of class sessions and used the same textbook and lab software.

The intervention design of the study was similar to previous studies that examined the association between computer-assisted instruction (CAI) and improved comprehension of conceptual relationships (Trumpower & Sarwar, 2010; Van der Merwe, 2012). The treatment group received access to short instructional videos and formative quizzes through an online tutorial site. The control group and treatment group completed the same statistical reasoning assessment.

Each instructor administered the online assessment during a regular class meeting after the ninth week of the semester when the textbook chapter about sampling distributions had been covered. The Assessment Resource Tools for Improving Statistical Thinking (ARTIST) scale regarding sampling variability served as the means of gauging student statistical reasoning abilities.

Research Questions and Null Hypotheses

The following research questions and null hypotheses were used to guide the study.

1. Is there a significant difference in mean ARTIST Scale score between students who were provided access to supplemental online instructional videos and formative quizzes and students who were not provided access?

    Ho1: There is no significant difference in mean ARTIST Scale score between the students who were provided access to supplemental online instructional videos and formative quizzes and students who were not provided access.

2. Is there a significant difference in mean ARTIST Scale score as compared by student demographics?

    a. Is there a significant difference in mean ARTIST Scale score as compared by age?

        Ho2a: There is no significant difference in mean ARTIST Scale score as compared by age.

b. Is there a significant difference in mean ARTIST Scale score between males and females?

Ho2b: There is no significant difference in mean ARTIST Scale score between males and females.

3. Is there a significant difference between mean ARTIST Scale scores when compared by number of formative quizzes successfully completed and average number of quiz attempts?

$Ho3_1$: There is no significant difference between mean ARTIST Scale scores when compared by number of formative quizzes successfully completed.

$Ho3_2$: There is no significant difference between mean ARTIST Scale scores when compared by average number of quiz attempts.

$Ho3_3$: There is no significant difference between mean ARTIST Scale scores when compared by a combination of number of formative quizzes successfully completed and average number of quiz attempts.

4. Is there a significant difference in mean ARTIST Scale score between students who opted not to access video tutorials and students who accessed the most video tutorials?

Ho4: There is no significant difference in mean ARTIST Scale score between students who opted not to access video tutorials and students who accessed the most video tutorials.

Instrumentation

An online statistical reasoning assessment was used as the primary instrument to collect data for this study. Similar assessments have been employed to determine the statistical reasoning ability of students in previous studies (Garfield, delMas, & Zieffler, 2012; Tintle, Topliff, Vanderstoep, Holmes, & Swanson, 2012). Additional data were gathered via a paper survey and from usage analytics from an online tutorial site. Instructors compiled all data into a single spreadsheet for each class.

The statistical reasoning assessment was administered online through the Assessment Resource Tools for Improving Statistical Thinking (ARTIST), a projected funded by the National Science Foundation (NSF CCLI –ASA- 0206571) and supported by the Foundation along with the University of Minnesota and California Polytechnic State University. The ARTIST Web site (https://apps3.cehd.umn.edu/artist/index.html) offers an overall Comprehensive Assessment of Outcomes in Statistics (CAOS) as well as 11 scales that focus on specific statistical topics (delMas et al., 2007). The ARTIST Scale dealing with sampling variability was chosen because the study's focus is limited to the particular difficulties students encounter when learning about sampling distributions.

The paper survey supplied instructions that the student enter first and last name as listed on the official student record. Providing this information ensured that instructors could link student demographic information with the ARTIST Scale scores. Other information requested on the survey included age and gender.

The tutorial site, Sophia.org, features short online videos and formative quizzes dealing with various academic subjects (Kessler, 2011). One of the full-time faculty members who taught a participating course in the study reviewed several tutorials and formative quizzes dealing with foundational topics in statistics and then created a closed group on the site. Students in the treatment group created free accounts and entered a code provided by their instructors to join the closed group. Instructors received access to a record of their students' level of activity on the site.

Instructors reported number of video tutorials accessed, number of formative quizzes successfully completed, and average number of quiz attempts. Previous studies have examined the relationship between total number of quizzes accessed and student achievement (Stull, Majerich et al., 2011; Wilson et al., 2011) as well as the relationship between the total number of multimedia objects accessed and student achievement (Bliwise, 2005; Miller, 2013).

Sample

Northeast State Community College is a higher education institution located in Blountville, Tennessee. As part of the state's college and university system, Northeast State is under the governance of the Tennessee Board of Regents. During the Fall 2013 semester, Northeast State enrolled 665 students in Probability & Statistics in 33 sections.

The population of this study consisted of 665 students who had registered for Probability & Statistics. Seven sections comprising a total of 190 students were taught by the four full-time faculty members who volunteered to participate in the study. The sample comprised 112 registered students who were still active in the seven course sections as of the twelfth week of a fifteen-week semester and who voluntarily completed the online statistical reasoning assessment during the appointed class meeting. Table 1 supplies sample demographic information.

Table 1

Sample Demographics

Variable		Frequency	%	Cumulative %
Gender	Female	72	64.3	64.3
	Male	40	35.7	100.0
Age	18-19	49	43.8	43.8
	20-22	30	26.8	70.6
	23 or older	33	29.4	100.0

Data Collection

The researcher applied for and obtained permission from the Institutional Review Board (IRB) at the home institution and from the IRB at Northeast State Community College to access data previously collected at Northeast State. Approval letters can be found in Appendix A. Students who opted to complete the ARTIST Sampling Variability Scale were provided with login information and an access code during a designated class meeting. The students completed a 15-question assessment. Additionally, the students submitted a brief paper survey to their respective instructors. The assessment scores were emailed to the instructors from the ARTIST project administrator.

With this information the instructors generated spreadsheets that omitted student names while listing the overall ARTIST scores, age, and gender. Instructors populated the spreadsheets with usage analytics from the Sophia.org tutorial site, including number of video tutorials accessed, number of formative quizzes successfully completed, and average number of quiz attempts.

Data Analysis

At the conclusion of the data collection phase, data were analyzed using Statistical Package for the Social Sciences (SPSS) data analysis software. Procedures for data analysis were guided by research questions used in the study. Independent variables were group membership (control or treatment), age range, gender, number of video tutorials accessed, number of formative quizzes successfully completed, and average number of quiz attempts. The ARTIST Scale score served as the dependent variable. Data were analyzed using analysis of variance (ANOVA) procedures (Green & Salkind, 2007). All data were analyzed at the .05 level of significance. Detailed results of each statistical procedure are provided in Chapter 4.

CHAPTER 4

FINDINGS

The purpose of this study was to determine whether significant differences in statistical reasoning abilities exist for completers of short online instructional videos and formative quizzes in undergraduate statistics courses. A series of research questions served as the guide for data analysis procedures. The independent variables included group membership (control or treatment), age range, gender, number of video tutorials accessed, number of formative quizzes successfully completed, and average number of quiz attempts. The ARTIST Scale score functioned as the dependent variable. The population for the study consisted of 665 undergraduate students enrolled in introductory statistics courses at a community college during the Fall 2013 semester.

This chapter provides a narrative of the demographic information for research participants and results from the research questions and related hypotheses. Analysis of variance (ANOVA) procedures were employed with an alpha level of .05 to establish significance in all the tests (Witte & Witte, 2010).

Demographics

Data for the study originated from responses to an online statistical reasoning assessment, usage statistics from an online tutorial site, and a brief demographic survey. The participants were undergraduate students enrolled in probability and statistics course during Fall 2013 semester. The study population consisted of 665 students who had enrolled in the probability and statistics course during the Fall 2013 semester. Using a cluster sampling strategy seven sections were selected for the study based upon the faculty members teaching the class. Of 190 total students who had enrolled in the selected sections, 112 voluntarily completed the online statistical reasoning assessment for a response rate of 59%. The sample comprised around 17% of the total population. Additionally, the sample consisted of 40 male (35.7%) and 72 female (64.3%) students. By age, 43.8% of participants were 18-19, 26.8% were 20-22, and 29.4% were age 23 or older.

Analyses of Research Questions

Four research questions were employed to guide the study, and seven corresponding null hypotheses were tested. Results of the statistical tests and related null hypotheses are presented in this section.

Research Question #1

> Is there a significant difference in mean ARTIST Scale score between students who were provided access to supplemental online instructional videos and formative quizzes and students who were not provided access?

> Ho1: There is no significant difference in mean ARTIST Scale score between students who were provided access to supplemental online instructional videos and formative quizzes and students who were not provided access.

An analysis of variance procedure was conducted to evaluate the difference in mean ARTIST Scale score between the students who were provided access to supplemental online instructional videos and formative quizzes (treatment group) and students who were not provided access (control group). The factor variable was group membership. The factor dependent variable was the ARTIST Scale score. The ANOVA was not significant, $F(1, 110) = .811$, $p = .370$. Therefore, the null hypothesis was retained.

Results indicate that students who were provided access to supplemental online instructional videos and formative quizzes (M=45.6) did not score significantly higher on the statistical reasoning assessment than students who were not provided access (M=43.4). The means and standard deviations for the groups are reported in Table 2. Figure 1 depicts the results graphically.

Table 2

Means and Standard Deviations of the Control and Treatment Groups

Group	*N*	*M*	*SD*
Control	51	43.4	14.0
Treatment	61	45.6	11.6

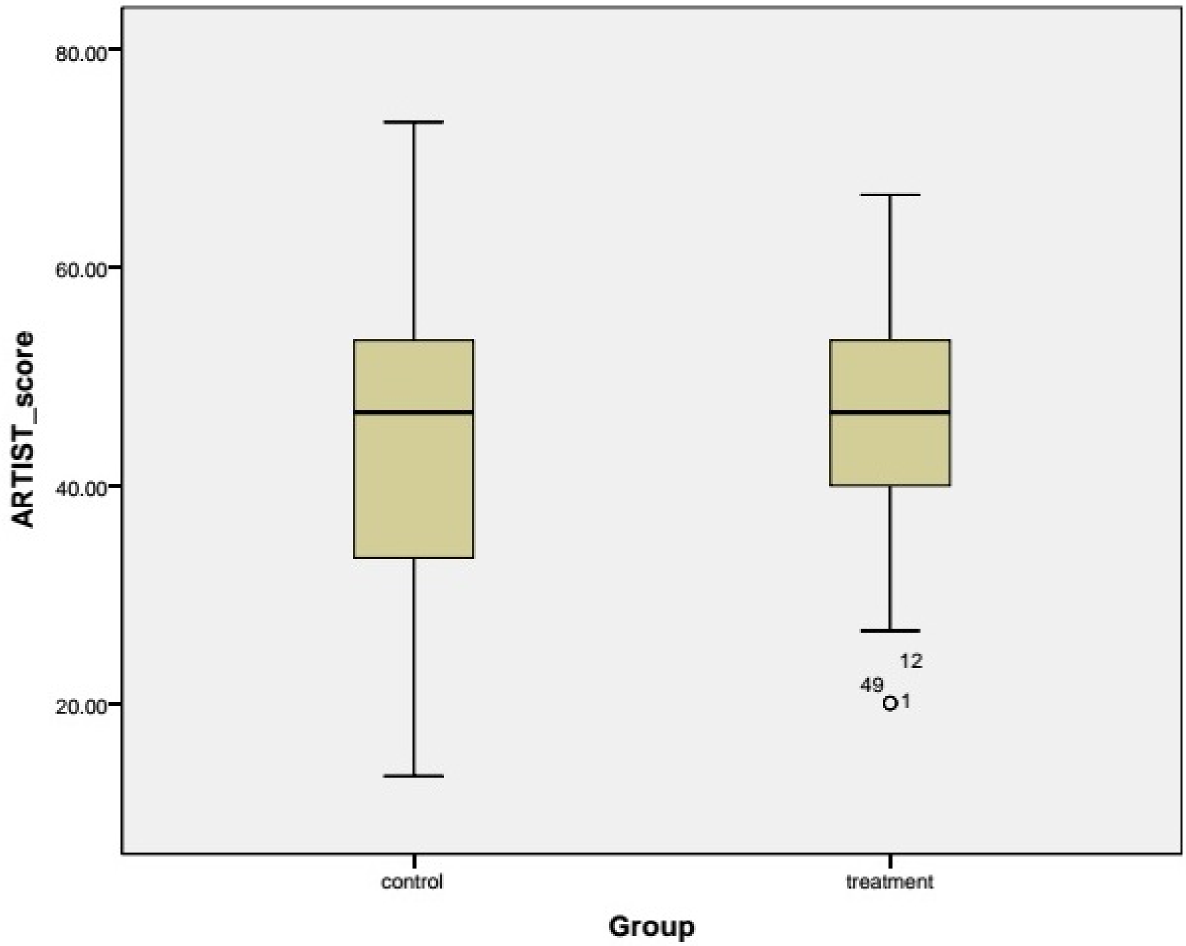


Figure 1. Boxplot of ARTIST Scale Score Based on Group

Research Question #2

Is there a significant difference in mean ARTIST Scale score as compared by student demographics?

Ho2a: There is no significant difference in mean ARTIST Scale score as compared by age.

A one-way analysis of variance was conducted to evaluate the difference in mean ARTIST Scale score between the three age groups. No significant differences exist between the three age groups, $F(2, 109) = 1.18$, $p = .312$. Therefore, the null hypothesis was retained. The effect size as measured by η^2, .021, was small.

Results indicate that scores on the statistical reasoning assessment did not differ significantly by age. The means and standard deviations for the groups are reported in Table 3. Figure 2 supplies a graphic representation of the results.

Table 3

Means and Standard Deviations of the Age Groups

Age Range	*N*	*M*	*SD*
18-19	49	46.5	12.3
20-22	30	44.0	12.9
23 or older	33	42.2	13.1

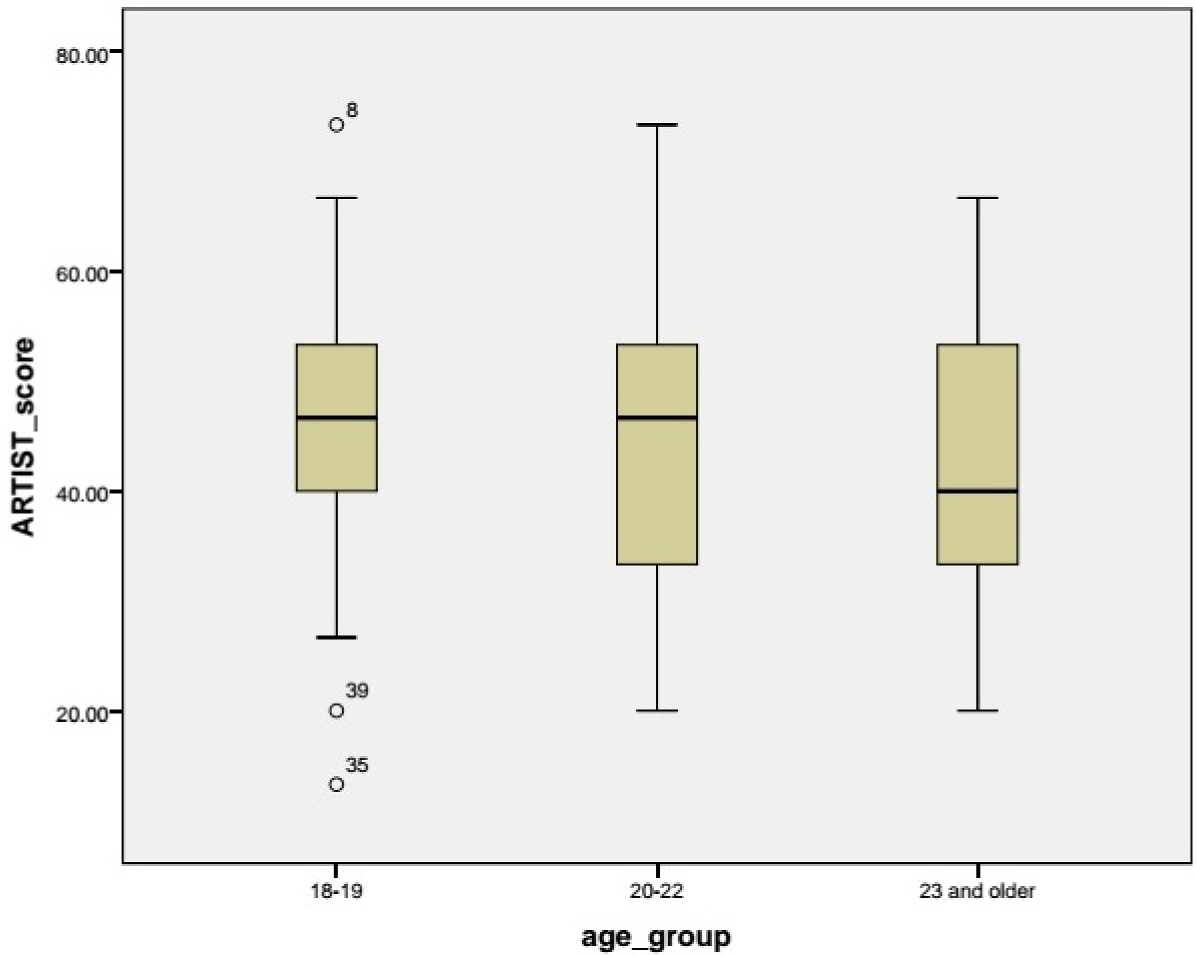


Figure 2. Boxplot of ARTIST Scale Score Based on Age

Ho2b: There is no significant difference in mean ARTIST Scale score between males and females.

An analysis of variance procedure was conducted to evaluate the difference in mean ARTIST Scale score between males and females. The factor variable was gender. The factor dependent variable was the ARTIST Scale score. The ANOVA was not significant, $F(1, 110) = .321$, $p = .572$. Therefore, the null hypothesis was retained.

Results indicate that scores on the statistical reasoning assessment did not differ significantly by gender. The means and standard deviations for the groups are reported in Table 4. Figure 3 supplies a graphic representation of the results.

Table 4

Means and Standard Deviations of Males and Females

Gender	*N*	*M*	*SD*
Male	40	43.7	13.6
Female	72	45.1	12.3

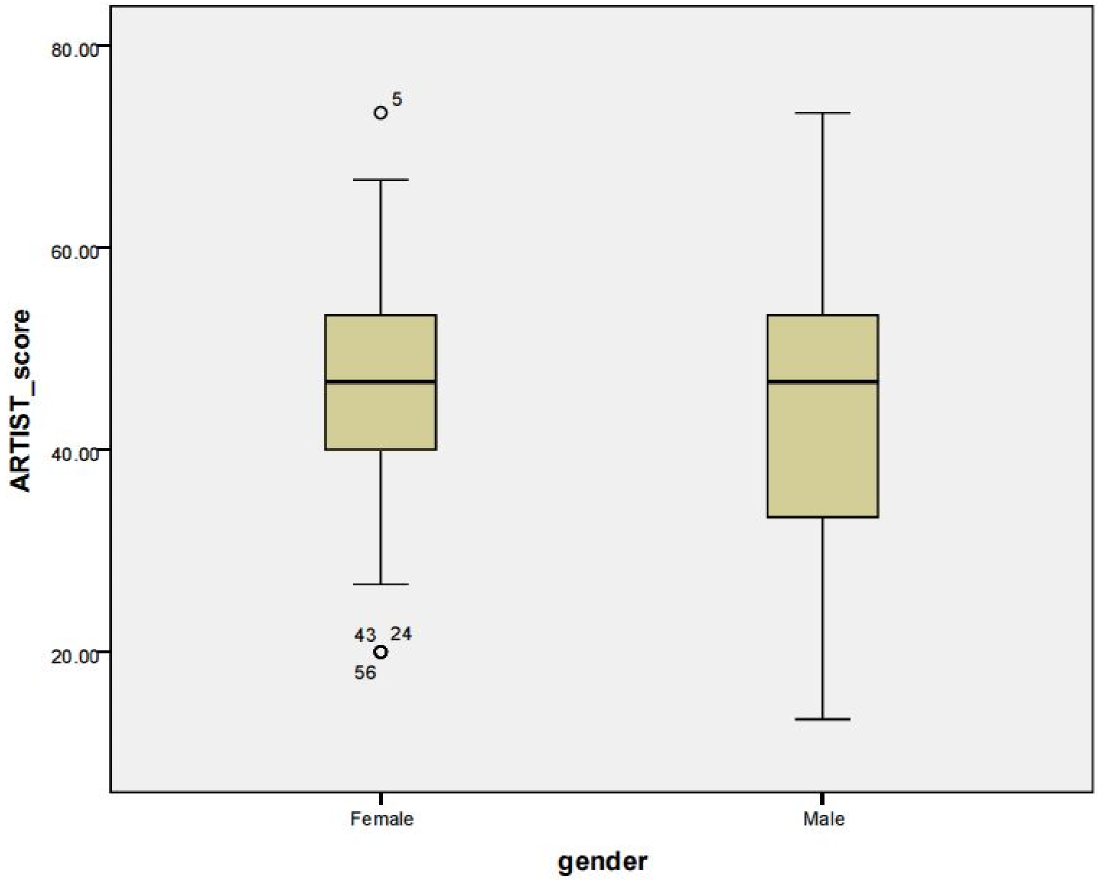


Figure 3. Boxplot of ARTIST Scale Score Based on Gender

Research Question #3

Is there a significant difference between mean ARTIST Scale scores when compared by number of formative quizzes successfully completed and average number of quiz attempts?

$Ho3_1$: There is no significant difference between mean ARTIST Scale scores when compared by number of formative quizzes successfully completed.

An analysis of variance procedure was conducted to evaluate the difference in mean ARTIST Scale scores between students who successfully completed seven formative quizzes and students who successfully completed seven or more formative quizzes. The midpoint number of seven quizzes was selected in order to allot an adequate number of participants to each group. The factor variable was the number of formative quizzes. The factor dependent variable was the ARTIST Scale score. The ANOVA was not significant, $F(1, 59) = .102$, $p = .751$. Therefore, the null hypothesis was retained.

Results indicate that students who successfully completed fewer than seven quizzes (M=46.9) did not score significantly higher on the statistical reasoning assessment than students who completed more than seven quizzes (M=44.2). The means and standard deviations for the groups are reported in Table 5. Figure 4 depicts the results graphically.

Table 5

Means and Standard Deviations of Quizzes Completed

Quizzes	*N*	*M*	*SD*
7 or more	31	44.2	12.1
Fewer than 7	30	46.9	11.1

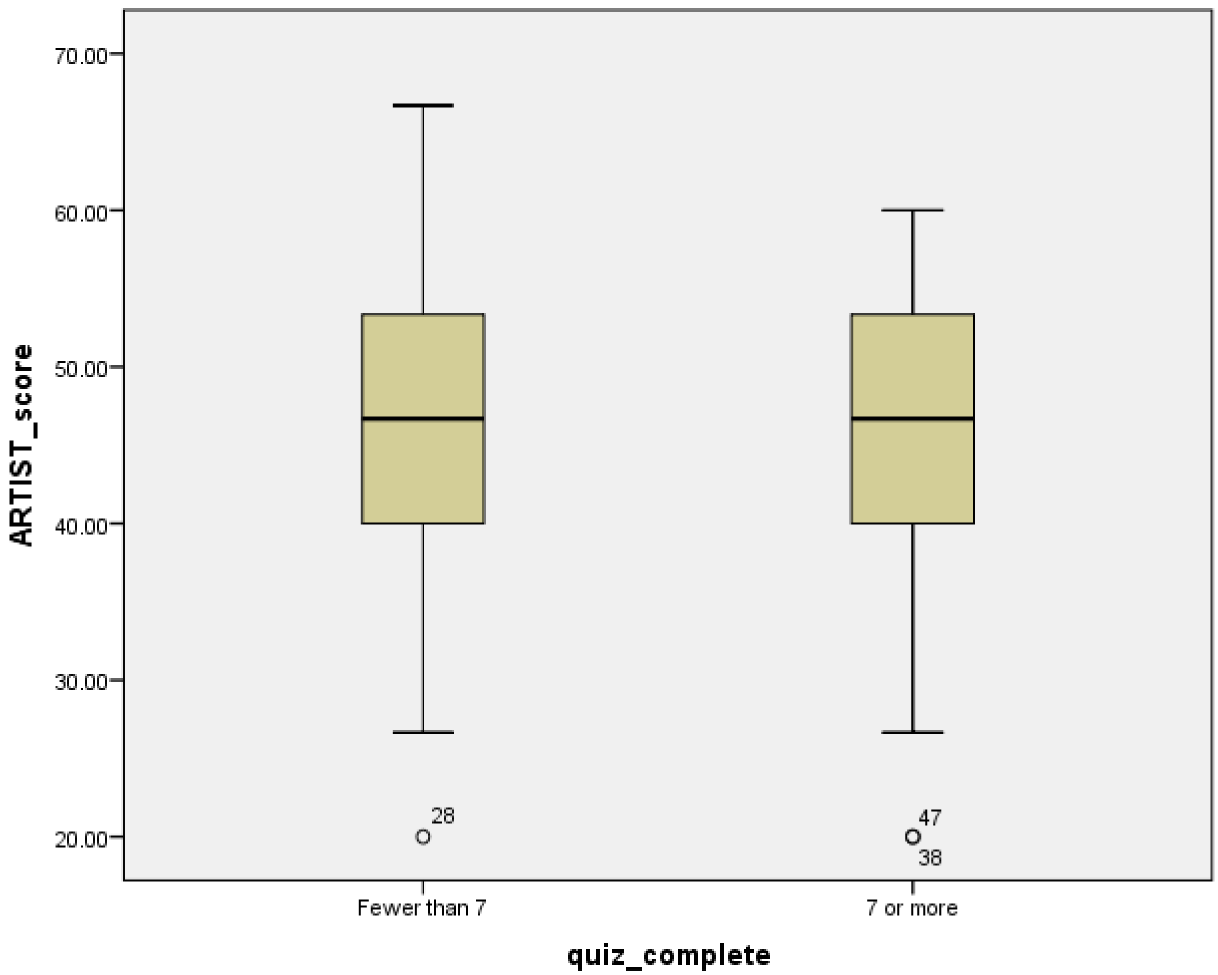


Figure 4. Boxplot of ARTIST Scale Score Based on Quizzes Completed

$Ho3_2$: There is no significant difference between mean ARTIST Scale scores when compared by average number of quiz attempts.

An analysis of variance procedure was conducted to evaluate the difference in mean ARTIST Scale scores between students who averaged 4.5 or more quiz attempts and students who averaged fewer than 4.5 attempts. The midpoint number of 4.5 attempts was selected in order to allot an adequate number of participants to each group. The factor variable was average number of quiz attempts. The factor dependent variable was the ARTIST Scale score. The ANOVA was not significant, $F(1, 59) = .364$, $p = .549$. Therefore, the null hypothesis was retained.

Results indicate that students who averaged fewer than 4.5 attempts per formative quiz (M=46.5) did not score significantly higher on the statistical reasoning assessment than students who averaged 4.5 attempts or more (M=45.3). The means and standard deviations for the groups are reported in Table 6. Figure 5 depicts the results graphically.

Table 6

Means and Standard Deviations of Average Quiz Attempts

Attempts	*N*	*M*	*SD*
4.5 or more	25	45.3	12.3
Fewer than 4.5	36	46.5	11.1

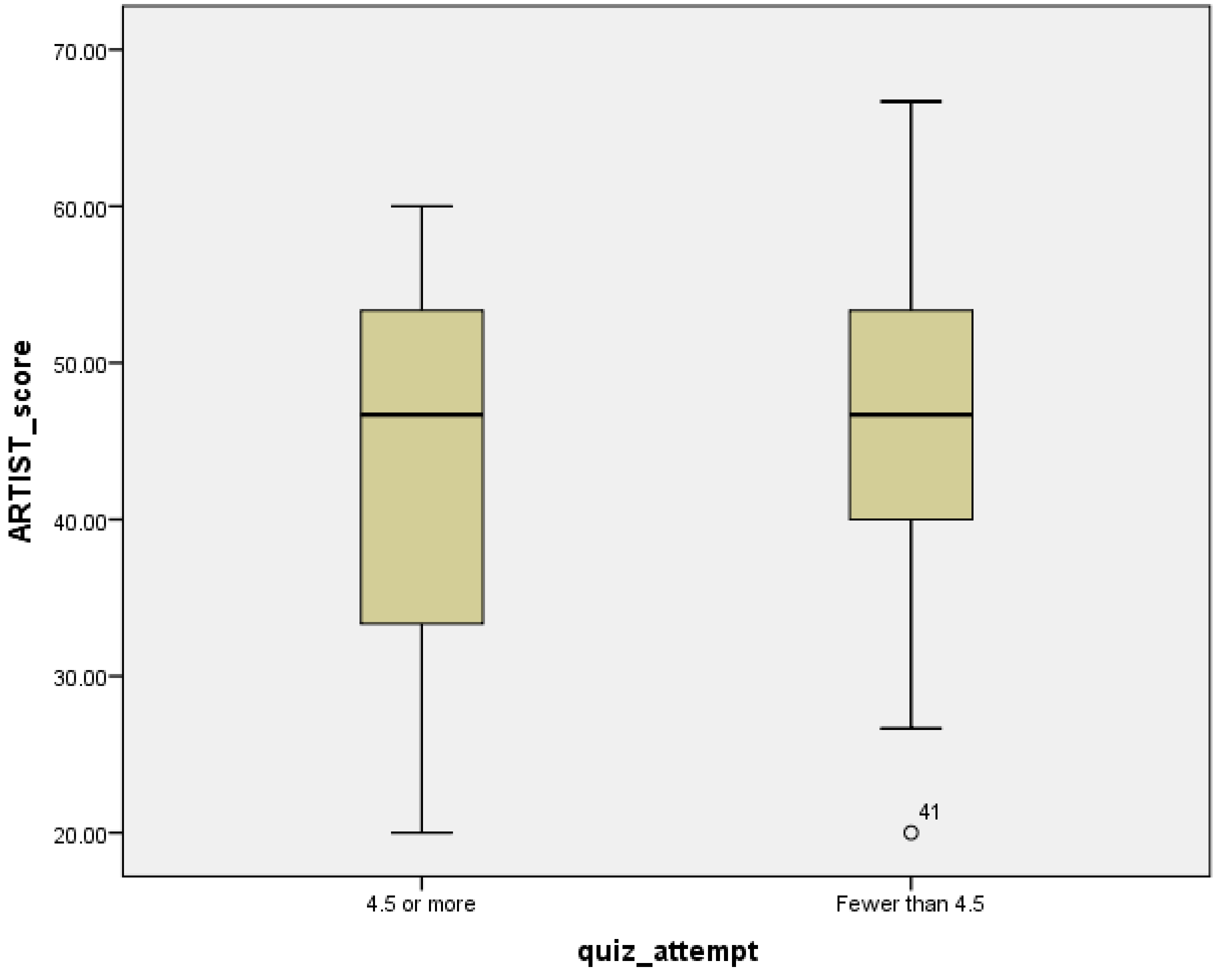


Figure 5. Boxplot of ARTIST Scale Score Based on Average Quiz Attempts

$Ho3_3$: There is no significant difference between mean ARTIST Scale scores when compared by a combination of number of formative quizzes successfully completed and average number of quiz attempts.

A two-way analysis of variance was conducted to evaluate the interaction between two factor variables. The factor variables were number of formative quizzes successfully completed and average number of quiz attempts. The factor dependent variable was the ARTIST Scale score. The main effect of number of quizzes was not significant, $F(1, 59) = .182$, $p = .671$. Similarly, the main effect of average quiz attempts was also not significant, $F(1, 59) = .369$, $p = .546$. Finally, the interaction between number of quizzes and average quiz attempts was not significant, $F(2, 58) = 1.54$, $p = .220$. Therefore, the null hypothesis was retained. The effect size as determined by η^2 was .003 for number of quizzes, and the effect size was .006 for average quiz attempts. The effect size for the interaction was .026.

Results indicate that the factor variables did not interact significantly in respect to the ARTIST Scale score. Additionally, analysis of main effects revealed no significant difference in means. The means and standard deviations for the factor variables are reported in Table 7. Figure 6 supplies a graphic representation of the results.

Table 7

Means and Standard Deviations of Quizzes Completed and Average Quiz Attempts

quiz_complete	quiz_attempt	N	M	SD
7 or more	4.5 or more	17	45.1	12.1
	Fewer than 4.5	13	43.1	12.4
Fewer than 7	4.5 or more	8	42.5	13.3
	Fewer than 4.5	23	48.4	10.2

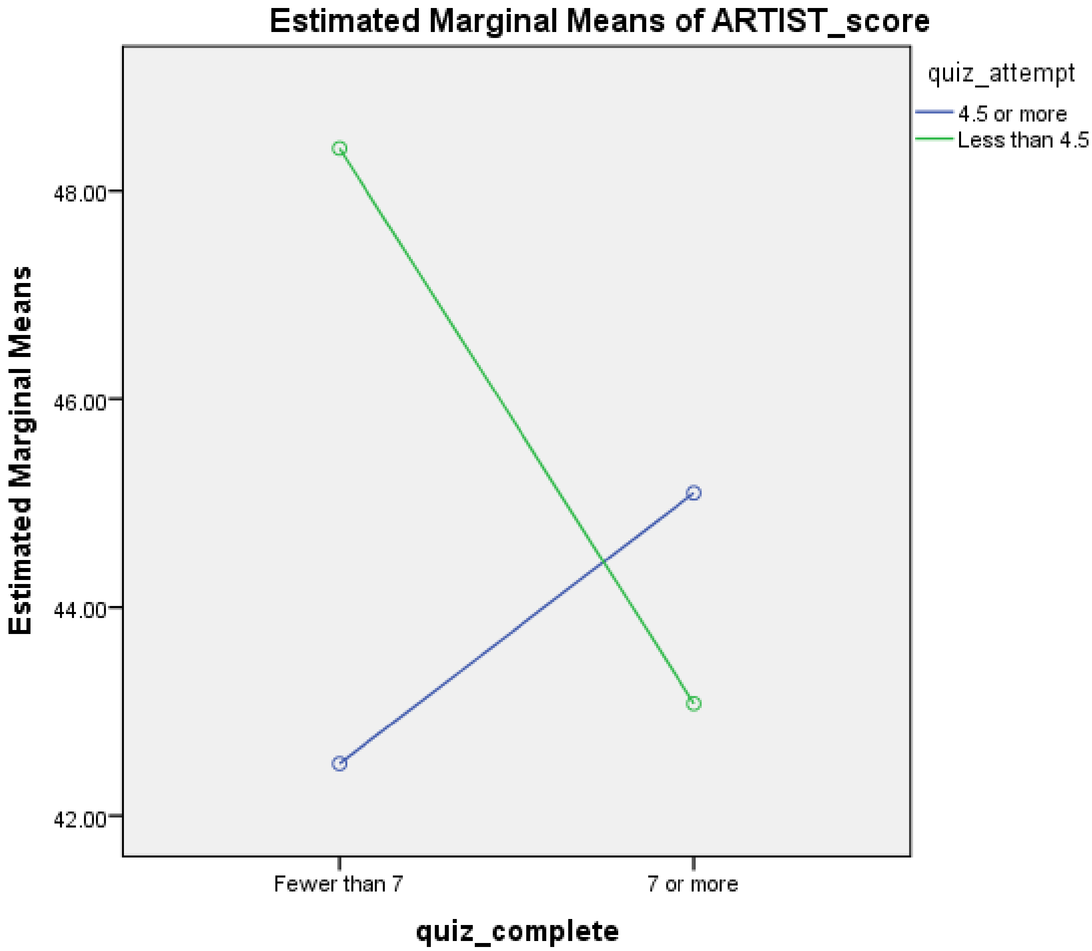


Figure 6. Estimated Marginal Means of ARTIST Scale Scores Based on Quizzes Completed and Average Quiz Attempts

Research Question #4

Is there a significant difference in mean ARTIST Scale score between students who opted not to access video tutorials and students who accessed the most video tutorials?

Ho4: There is no significant difference in mean ARTIST Scale score between students who opted not to access video tutorials and students who accessed the most video tutorials.

An analysis of variance procedure was conducted to evaluate the difference in mean ARTIST Scale score between students who opted not to access video tutorials and students who accessed the most video tutorials. The factor variable was the number of tutorials accessed. The factor dependent variable was the ARTIST Scale score. The ANOVA was not significant, $F(1, 39) = .425$, $p = .518$. Therefore, the null hypothesis was retained.

Results indicated that scores on the statistical reasoning assessment did not differ significantly between students who accessed the most video tutorials (M=46.0) and the students who opted not to access video tutorials (M=48.3).

Table 8

Means and Standard Deviations of the Number of Video Tutorials Accessed

Videos accessed	*N*	*M*	*SD*
Most videos	20	46.0	11.6
No video	20	48.3	11.0

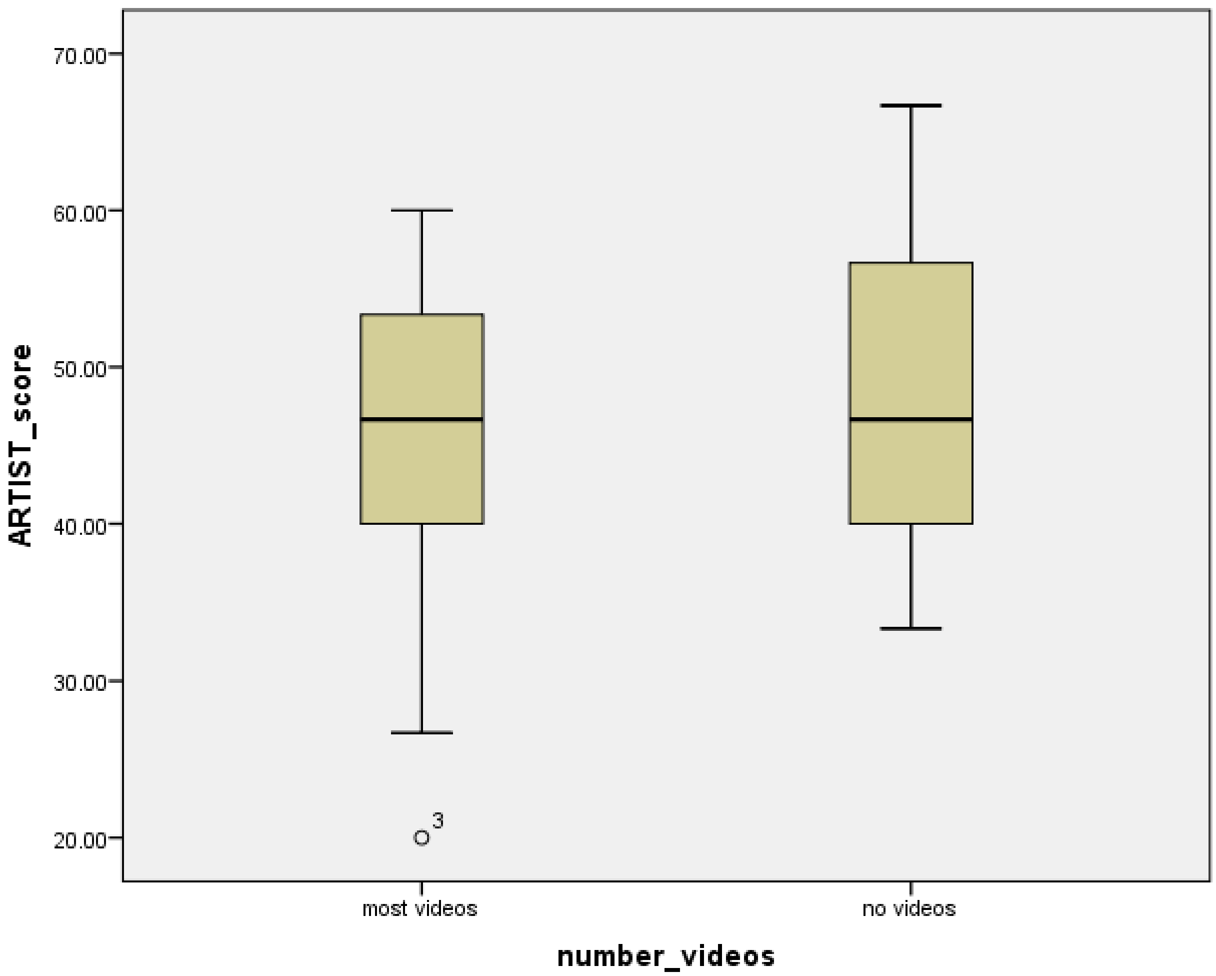


Figure 7. Box Plot of ARTIST Scale Score Based on Number of Video Tutorials Accessed

Chapter Summary

This chapter reviewed the statistical analyses of data obtained from an online statistical reasoning assessment, usage statistics at an online tutorial site, and a brief demographic survey. Four research questions and seven hypotheses guided the analysis procedures. All data were collected from undergraduate students enrolled in seven sections of a probability and statistics course during the Fall 2013 semester. There were 112 students who completed the assessment and survey. The data were analyzed using ANOVA statistical procedures, and results were displayed via tables and graphs. Chapter 5 contains a summary of findings, conclusions, implications for practice, and recommendations for future research.

CHAPTER 5

SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

Chapter 5 presents a summary of findings, conclusions, and recommendations for future research concerning the combination of multimedia tutorials and online formative quizzes for resolving undergraduate learner misconceptions. The findings are presented in terms of the research questions that guided the study. The purpose of this study was to determine whether significant differences in statistical reasoning abilities a relationship exist for completers of short online instructional videos and formative quizzes in undergraduate statistics courses and by demographic variables.

Summary of Findings

The data were collected from 40 male (35.7%) and 72 female (64.3%) students. This proportion conforms to current trends for community colleges in the United States, where female student enrollment generally exceeds male student enrollment (Horn, Nevill, & Griffith, 2006). Moreover, 43.8% of participants were aged 18-19, 26.8% were 20-22, and 29.4% were 23 or older. Four research questions with seven null hypotheses guided data collection and analysis. These hypotheses were tested using analysis of variance (ANOVA) procedures, and results were presented in Chapter 4. An alpha of .05 served as the parameter for establishing significance in all procedures.

Research Question #1

Is there a significant difference in mean ARTIST Scale score between students who received access to supplemental online instructional videos and formative quizzes and students who did not receive access?

The ANOVA did not reveal a significant difference in mean scores. ARTIST Scale score had served as an operationalization for statistical reasoning ability, and therefore results indicate that the statistical reasoning ability of students who were provided access to supplemental online videos and formative quizzes did not differ significantly from students who were not provided access.

Previous studies have demonstrated significant links between computer-assisted instruction and conceptual reasoning ability (Trumpower & Sarwar, 2010; Van der Merwe, 2012). Other studies have shown significant connections between online formative assessment and student achievement (Davis, 2013; Limniou & Smith, 2014; Stull et al., 2011; Wilson et al., 2011) and between multimedia interventions and student achievement (Bastürk, 2005; Bliwise, 2005; Miller, 2013). These groups of studies provided a promising basis for exploring connections between supplemental online videos and formative quizzes and statistical reasoning ability.

Research Question #2

Is there a significant difference in mean ARTIST Scale score as compared by student demographics?

Neither age range nor gender proved to be a significant factor for influencing statistical reasoning ability. The ANOVA procedures revealed no significant difference in mean score between the three age ranges nor between male and female. These results indicate that statistical reasoning ability does not significantly differ by age range or gender.

Research Question #3

Is there a significant difference in mean ARTIST Scale score when compared by number of formative quizzes successfully completed and average number of quiz attempts?

This question addressed the relationship between ARTIST Scale scores and individual factors as well as the level of interaction between factors in regard to ARTIST scores. Neither of the individual factors alone demonstrated a significant difference in ARTIST scores. Likewise, the factor variables did not interact significantly in respect to the ARTIST scores. Other studies featuring optional quizzes for complex academic subjects had supplied a basis for expecting a significant interaction, although in terms of exam grades instead of statistical reasoning ability (Brothen & Wambach, 2001; Johnson, 2006; McKeown & Maclean, 2013; Scott et al., 2014).

Research Question #4

Is there a significant difference in mean ARTIST Scale score between students who opted not to access video tutorials and students who accessed the most video tutorials?

The ANOVA did not reveal a significant difference in mean scores. Therefore, results indicate that the statistical reasoning ability of students who accessed the most video tutorials did not differ significantly from students who opted not to access video tutorials. Previous studies had demonstrated significant links between multimedia interventions and student achievement and supported the expectation of similar findings in the present study (Bastürk, 2005; Bliwise, 2005; Miller, 2013).

Recommendations for Policy

The following policy recommendations may be derived from the study.

1. It is recommended that academic divisions and departments review and consider the existing practices and purposes for using supplemental instructional videos as course elements.
2. Moreover, it is recommended that departments deploy course analytics to evaluate learner engagement and support other forms of assessment (Kizilcec, Piech, & Schneider, 2013; Lockyer, Heathcote, & Dawson, 2013).

Recommendations for Practice

1. Based upon parameters of the study, undergraduate students did not benefit significantly from receiving access to supplemental videos and formative quizzes although there is a great deal of evidenced empirical support for such practices. It is therefore recommended that instructors consider the unique learning outcomes

proceed with caution before implementing a combination of online video tutorials and quizzes as part of an introductory statistics course.

2. Additionally, it is recommended that instructors provide substantial feedback when incorporating formative quizzes as part of a course (Limniou & Smith, 2014; G. Smith, 2007). Examples of substantial feedback include detailed descriptions of why a particular answer is incorrect or a review of the concept to help learners discover their mistakes.

Recommendations for Future Research

The following recommendations are offered as possible opportunities for research in the same area or topic:

1. This study addressed the use of supplemental online video tutorials and formative quizzes in an introductory statistics course. Future research could examine the benefits of online videos and quizzes when integrated as essential course elements.
2. Additionally, this study focused on a single institution in the southeastern United States. Further studies could involve a larger sample drawn from multiple undergraduate institutions from different regions as well as from other countries.
3. Finally, data for this study were drawn from a single academic term in the disciplinary area of statistics. Future studies could examine and compare samples from multiple terms, academic years, and disciplines.

Conclusions

In conclusion, the statistical reasoning abilities of undergraduate students at a community college did not differ significantly based upon access to supplemental online instructional videos and formative quizzes. Additionally, statistical reasoning abilities did not differ significantly between males and females or among age ranges.

Moreover, statistical reasoning abilities did not differ significantly in regard to number of formative quizzes successfully completed or average number of quiz attempts. Furthermore, statistical reasoning abilities did not differ significantly in regard to number of video tutorials accessed.

REFERENCES

Aberson, C. L., Berger, D. E., Healy, M. R., & Romero, V. L. (2003). Evaluation of an interactive tutorial for teaching hypothesis testing concepts. *Teaching of Psychology, 30*(1), 75-78. doi: 10.1207/S15328023TOP3001_12

Aberson, C. L., Berger, D. E., Healy, M. R., Kyle, D. J., & Romero, V. L. (2000). Evaluation of an interactive tutorial for teaching the central limit theorem. *Teaching of Psychology, 27*(4), 289-291. doi: 10.1207/S15328023TOP2704_08

Anderson, R. C., Spiro, R. J., & Anderson, M. C. (1978). Schemata as scaffolding for the representation of information in connected discourse. *American Educational Research Journal, 15*(3), 433-440. doi: 10.2307/1162496

Baddeley, A. D. (1986). *Working memory*. Oxford, England, UK: Clarendon.

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford, England, UK: Oxford University Press.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29. doi: 10.1146/annurev-psych-120710-100422

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

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development, 67*(3), 1206-1222.

Bartlett, F. C. (1932). A theory of remembering. In *Remembering: A study in experimental and social psychology* (pp. 197-214). Cambridge, England, UK: Cambridge University Press.

Bastürk, R. (2005). The effectiveness of computer-assisted instruction in teaching introductory statistics. *Educational Technology & Society, 8*(2), 170-178. Retrieved October 12, 2013, from http://www.ifets.info/journals/8_2/16.pdf

Batanero, C., Tauber, L. M., & Sánchez, V. (2004). Students' reasoning about the normal distribution. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 257-276). Dordrecht, The Netherlands: Kluwer Academic.

Ben-Zvi, D., & Garfield, J. B. (2004). *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht, The Netherlands: Kluwer Academic.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London, UK: School of Education, King's College London.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. doi: 10.1007/s11092-008-9068-5

Bliwise, N. G. (2005). Web-based tutorials for teaching introductory statistics. *Journal of Educational Computing Research, 33*(3), 309-325. doi: 10.2190/0D1J-1CE1-5UXY-3V34

Block, J. H. (1977). Individualized instruction: A mastery learning perspective. *Educational Leadership, 34*(5), 337-341.

Block, J. H. (1980). Promoting excellence through mastery learning. *Theory Into Practice, 19*(1), 66-74. doi: 10.1080/00405848009542874

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 1-12.

Bloom, B. S. (1971). Mastery learning. In J. H. Block (Ed.), *Mastery learning: Theory and practice* (pp. 47-63). New York, NY: Holt, Rinehart and Winston.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16. doi: 10.3102/0013189X013006004

Borman, G. D. (2002). Experiments for educational evaluation and improvement. *Peabody Journal of Education, 77*(4), 7-27.

Brothen, T., & Wambach, C. (2001). Effective student use of computerized quizzes. *Teaching of Psychology, 28*(4), 292-294. doi:10.1207/S15328023TOP2804_10

Brown, D. E. (1992). Using examples and analogies to remediate misconceptions in physics: Factors influencing conceptual change. *Journal of Research in Science Teaching, 29*(1), 17-34. doi: 10.1002/tea.3660290104

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1999). Sources of conceptual change. In E. K. Schlonick, K. Nelson, S. A. Gelman, & P. H. Miller (Eds.), *Conceptual development: Piaget's legacy* (pp. 293-326). Mahwah, NJ: Erlbaum.

Castro Sotos, A., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2*(2), 98-113. doi: 10.1016/j.edurev.2007.04.001

Chance, B., delMas, R., & Garfield, J. B. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295-323). Dordrecht, The Netherlands: Kluwer Academic.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332. doi: 10.1207/s1532690xci0804_2

Chi, M. H., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2012). Misconceived causal explanations for emergent processes. *Cognitive Science, 36*, 1-61. doi: 10.1111/j.1551-6709.2011.01207.x

Chi, M. T., & Hausmann, R. G. (2003). Do radical discoveries require ontological shifts? In L. V. Shavinina & R. Sternberg (Eds.), *The international handbook on innovation* (pp. 430-444). Oxford, England, UK: Elsevier Science.

Chi, M. T., & Roscoe, R. D. (2002). The processes and challenges of conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 3-27). Dordrecht, The Netherlands: Kluwer Academic.

Chi, M. T. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14*(2), 161-199. doi: 10.1207/s15327809jls1402_1

Chi, M. T. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 61-82). Hillsdale, NJ: Erlbaum.

Chi, M. T., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction, 4*(1), 27-43. doi: 10.1016/0959-4752(94)90017-5

Christensen, C., & Weise, M. R. (2014, May 9). MOOCs' disruption is only beginning. The Boston Globe. Retrieved September 7, 2014, from http://www.bostonglobe.com/opinion/2014/05/09/moocs-disruption-only-beginning/S2VlsXpK6rzRx4DMrS4ADM/story.html

Clark, R. C., & Mayer, R. E. (2011). *E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (3rd ed.). San Francisco, CA: Pfeiffer.

Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), 1-15.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Davis, K. A. (2013). Using low- and no-stakes quizzing for student self-evaluation of readiness for exams. *International Journal of Construction Education and Research, 9*(4), 256-271. doi: 10.1080/15578771.2013.809036

delMas, R. C., Garfield, J. B., Ooms, A., & Chance, B. L. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28-58. Retrieved October 12, 2013, from https://www.stat.auckland.ac.nz/~iase/serj/SERJ6(2)_delMas.pdf diSessa, A. (1988). Knowledge in pieces. In G. E. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age* (pp. 49-70). Hillsdale, NJ: Erlbaum.

diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction, 10*(2-3), 105-225. doi: 10.1080/07370008.1985.9649008

diSessa, A. A. (2002). Why "conceptual ecology" is a good idea (M. Limón & L. Mason, Eds.). In *Reconsidering conceptual change: Issues in theory and practice* (pp. 28-60). Dordrecht, The Netherlands: Kluwer Academic.

diSessa, A., Gillespie, N., & Esterly, J. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*(6), 843-900. doi: 10.1016/j.cogsci.2004.05.003

Driver, R., & Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education, 5*(1), 61-84. doi: 10.1080/03057267808559857

Fischer, K. (2013, March 12). A college degree sorts job applicants, but employers wish it meant more. *The Chronicle of Higher Education*. Retrieved October 12, 2013, from http://chronicle.com/article/The-Employment-Mismatch/137625/

Gal, I., & Garfield, J. B. (1997). *The assessment challenge in statistics education*. Amsterdam, The Netherlands: IOS Press.

Gambari, A. I., Falode, C. O., & Adegbenro, D. A. (2014). Effectiveness of computer animation and geometrical instructional model on mathematics achievement and retention among junior secondary school students. *European Journal of Science and Mathematics*

Education, 2(2), 127-146. Retrieved April 22, 2014, from http://www.scimath.net/articles/22/226.pdf

Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal, 2*(1), 22-38. Retrieved October 12, 2013, from http://iase-web.org/documents/SERJ/SERJ2(1).pdf

Garfield, J., DelMas, R., & Zieffler, A. (2010). Assessing important learning outcomes in introductory tertiary statistics courses. In P. Bidgood, N. Hunt, & F. R. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 75-86). Chichester, West Sussex, U.K.: John Wiley.

Gore, P. A. (2006). Academic self-efficacy as a predictor of college outcomes: Two incremental validity studies. *Journal of Career Assessment, 14*(1), 92-115. doi: 10.1177/1069072705281367

Green, S. B., & Salkind, N. J. (2007). *Using SPSS for Windows and Macintosh: Analyzing and understanding data*. Upper Saddle River, NJ: Prentice Hall.

Guskey, T. R. (1997). *Implementing mastery learning* (2nd ed.). Belmont, CA: Wadsworth.

Guskey, T. R. (2007). Closing achievement gaps: Revisiting Benjamin S. Bloom's "learning for mastery" *Journal of Advanced Academics, 19*(1), 8-31. doi: 10.4219/jaa-2007-704

Guskey, T. R. (2010). Lessons of mastery learning. *Educational Leadership, 68*(2), 52-7. Retrieved September 7, 2014, from http://www.ascd.org/publications/educational-leadership/oct10/vol68/num02/Lessons-of-Mastery-Learning.aspx

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. doi: 10.3102/003465430298487

Horn, L., Nevill, S., & Griffith, J. (2006). *Profile of undergraduates in U.S. postsecondary education institutions, 2003-04: With a special analysis of community college students* (Rep. No. NCES 2006-184). Washington, DC: National Center for Education Statistics. (ERIC Document Reproduction Service No. ED491908)

Johnson, G. M. (2006). Optional online quizzes: College student use and relationship to achievement. *Canadian Journal of Learning and Technology, 32*(1). Retrieved October 12, 2013, from http://cjlt.csj.ualberta.ca/index.php/cjlt/article/view/61/58

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

Kant, I. (1884). *Critique of pure reason* (J. M. Meiklejohn, Trans.). London: Bell.

Kessler, S. (2011, March 9). New platform uses social features to improve online education. Retrieved October 24, 2013, from http://mashable.com/2011/03/09/sophia/

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. Third International Conference on Learning Analytics and Knowledge. Retrieved June 11, 2013, from http://rene.kizilcec.com/

Koller, D. (2012, June 26). *What we're learning from online education*. Lecture presented at TEDGlobal 2012 in Edinburgh International Conference Center, Edinburgh, Scotland, UK. Retrieved February 1, 2014, from http://www.ted.com/talks/daphne_koller_what_we_re_learning_from_online_education

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Kulik, C. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*(2), 265-299. doi: 10.3102/00346543060002265

Larwin, K., & Larwin, D. (2011). A meta-analysis examining the impact of computer-assisted instruction on postsecondary statistics education: 40 years of research. *Journal of Research on Technology in Education, 43*(3), 253-278.

Lawton, D., Vye, N., Bransford, J., Sanders, E., Richey, M., French, D., & Stephens, R. (2012). Online learning based on essential concepts and formative assessment. *Journal of Engineering Education, 101*(2), 244-287. doi: 10.1002/j.2168-9830.2012.tb00050.x

Lewin, T. (2013, January 6). Students rush to web classes, but profits may be much later. *The New York Times*. Retrieved October 2, 2014, from http://www.nytimes.com/2013/01/07/education/massive-open-online-courses-prove-popular-if-not-lucrative-yet.html

Limniou, M., & Smith, M. (2014). The role of feedback in e-assessments for engineering education. *Education and Information Technologies, 19*(1), 209-225.

Lloyd, S. A., & Robertson, C. L. (2012). Screencast tutorials enhance student learning of statistics. *Teaching of Psychology, 39*(1), 67-71. doi: 10.1177/0098628311430640

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. American Behavioral Scientist, 57(10), 1439-1459. doi:10.1177/0002764213479367

Maguire Associates. (2012, December). *The role of higher education in career development: Employer perceptions* (Rep.). Retrieved September 7, 2014, from http://chronicle.com/items/biz/pdf/Employers%20Survey.pdf

Marshall, S. P. (1995). *Schemas in problem solving*. Cambridge, England: Cambridge University Press.

Matthews, D. (2013, September 5). The tuition is too damn high, part IX: Will MOOCs save us? *The Washington Post*. Retrieved February 2, 2015, from http://www.washingtonpost.com/blogs/wonkblog/wp/2013/09/05/the-tuition-is-too-damn-high-part-ix-will-moocs-save-us/

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, England, UK: Cambridge University Press.

Mayer, R. E. (2002). Understanding conceptual change: A commentary. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 101-111). Dordrecht, The Netherlands: Kluwer Academic.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In *The Cambridge handbook of multimedia learning* (pp. 31-48). New York, NY: Cambridge University Press.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*(8), 760-769. doi: 10.1037/0003-066X.63.8.760

Mckeown, P., & Maclean, G. (2013). Is activity in online quizzes correlated with higher exam marks? *New Zealand Economic Papers, 47*(3), 276-287. doi:10.1080/00779954.2012.715826

Merriënboer, J. J., & Kirschner, P. A. (2012). *Ten steps to complex learning: A systematic approach to four-component instructional design* (2nd ed.). New York, NY: Routledge.

Miller, N. B. (2013). Student access of supplemental multimedia and success in an online course. *American Journal of Distance Education, 27*(4), 242-252. doi: 10.1080/08923647.2013.837727

NACE. (2013, November). *Job Outlook 2014* (Rep.). Retrieved September 7, 2014, from https://www.naceweb.org/s10022013/job-outlook-skills-quality.aspx

Noer, M. (2012, November 2). One man, one computer, 10 million students: How Khan Academy Is reinventing education. *Forbes*. Retrieved October 12, 2013, from http://www.forbes.com/sites/michaelnoer/2012/11/02/one-man-one-computer-10-million-students-how-khan-academy-is-reinventing-education/

Novick, S., & Nussbaum, J. (1981). Pupils' understanding of the particulate nature of matter: A cross-age study. *Science Education, 65*(2), 187-196. doi: 10.1002/sce.3730650209

Paivio, A. (1971). *Imagery and verbal processes.* New York, NY: Holt, Rinehart and Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

Paivio, A. (2006). *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum.

Piaget, J., & García, R. (1974). *Understanding causality* (D. Miles, Trans.). New York, NY: Norton.

Piaget, J. (1926). *The language and thought of the child*. New York, NY: Harcourt.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211-227.

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science, 28*(1), 4-13. doi: 10.1002/bs.3830280103

Reading, C., & Shaughnessy, M. (2004). Reasoning about variation. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 201-226). Dordrecht: Kluwer Academic.

Rumelhart, D. E., & Norman, D. A. (1976). *Accretion, tuning, and restructuring: Three modes of learning* (Rep. No. 7802). Retrieved November 10, 2013 from http://files.eric.ed.gov/fulltext/ED134902.pdf (ERIC Document Reproduction Service No. ED134902)

Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257-270. doi: 10.1023/A:1023692604014

Scott, T. E., Thomas, L., Edwards, K., Jones, J., Swan, H., & Wessels, A. (2014). Increasing recognition of fetal heart anatomy using online tutorials and mastery learning compared with classroom instructional methods. *Journal of Diagnostic Medical Sonography, 30*(2), 52-59. doi:10.1177/8756479313517297

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 1, AERA Monograph Series -- Curriculum Evaluation, pp. 39-83). Chicago, IL: Rand McNally.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making, 10*(1), 33-51. doi: 10.1002/(SICI)1099-0771(199703)10:13.0.CO;2-6

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shi, L. (2012). *The effects of a custom-designed animation program on learning Chinese characters* (Unpublished doctoral dissertation). George Mason University. Retrieved April 22, 2014, from http://digilib.gmu.edu/xmlui/handle/1920/7935

Slavin, R. E. (2008). Perspectives on evidence-based research in education--what works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5-14.

Slotta, J. D., Chi, M. T., & Joram, E. (1995). Assessing students' misclassifications of physics concepts: An ontological basis for conceptual change. *Cognition and Instruction, 13*(3), 373-400. doi: 10.1207/s1532690xci1303_2

Smith, C. (2007). Bootstrapping processes in the development of students' commonsense matter theories: Using analogical mappings, thought experiments, and learning to measure to promote conceptual restructuring. *Cognition and Instruction, 25*(4), 337-398. doi: 10.1080/07370000701632363

Smith, G. (2007). How does student performance on formative assessments relate to learning assessed by exams? *Journal of College Science Teaching, 36*(7), 28-34. Retrieved October 12, 2013, from http://learningcenter.nsta.org/product_detail.aspx?id=10.2505/4/jcst07_036_07_28

Sosa, G. W., Berger, D. E., Saw, A. T., & Mary, J. C. (2011). Effectiveness of computer-assisted instruction in statistics: A meta-analysis. *Review of Educational Research, 81*(1), 97-128. doi: 10.3102/0034654310378174

Stull, J. C., Majerich, D. M., Bernacki, M. L., Varnum, S. J., & Ducette, J. P. (2011). The effects of formative assessment pre-lecture online chapter quizzes and student-initiated inquiries to the instructor on academic achievement. *Educational Research and Evaluation, 17*(4), 253-262. doi: 10.1080/13803611.2011.621756

Swartz, A. (2015, February 19). Khan Academy program wants Bay Area kids to get interested in math. *Silicon Valley Business Journal*. Retrieved February 22, 2015, from http://www.bizjournals.com/sanjose/news/2015/02/19/khan-academy-program-wants-bay-area-kids-to-get.html

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295-312. doi: 10.1016/0959-4752(94)90003-5

Talbert, P. (2012). Strategies to increase enrollment, retention, and graduation rates. *Journal of Developmental Education, 36*(1), 22-36.

Tintle, N., Topliff, K., Vanderstoep, J., Holmes, V., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomized-based introductory statistics curriculum. *Statistics Education Research Journal, 11*(1), 21-40.

Trochim, W., & Donnelly, J. (2006). *The research methods knowledge base*. Stamford, CT: Cengage Learning.

Trumpower, D. L., & Sarwar, G. S. (2010). Effectiveness of structural feedback provided by Pathfinder networks. *Journal of Educational Computing Research, 43*(1), 7-24. doi: 10.2190/EC.43.1.b

Tugend, A. (2013, June 29). What it takes to make new college graduates employable. *The New York Times*. Retrieved October 12, 2013 from http://www.nytimes.com/2013/06/29/your-money/a-quest-to-make-college-graduates-employable.html

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*(2), 105-110. doi: 10.1037/h0031322

Udell, J. (2004). *Name that genre: Screencast*. Retrieved November 10, 2013, from http://jonudell.net/udell/2004-11-17-name-that-genre-screencast.html

Van der Merwe, L. (2012). *An analysis of the field of statistics education research with special reference to computer-assisted teaching and learning* (Unpublished doctoral dissertation). University of the Free State, Bloemfontein, ZA.

Vosniadou, S., & Verschaffel, L. (2004). Extending the conceptual change approach to mathematics learning and teaching. *Learning and Instruction, 14*(5), 445-451. doi: 10.1016/j.learninstruc.2004.06.014

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 45-69.

Vosniadou, S. (2002). On the nature of naïve physics. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 61-76). Dordrecht, The Netherlands: Kluwer Academic.

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 3-34). New York, NY: Routledge.

Wilson, K., Boyd, C., Chen, L., & Jamal, S. (2011). Improving student performance in a first-year geography course: Examining the importance of computer-assisted formative assessment. *Computers & Education, 57*(2), 1493-1500. doi: 10.1016/j.compedu.2011.02.011

Witte, R. S., & Witte, J. S. (2010). *Statistics*. Hoboken, NJ: Wiley.

Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist, 27*(4), 531-541. doi: 10.1207/s15326985ep2704_8

Özdemir, G., & Clark, D. B. (2007). An overview of conceptual change theories. *Eurasia Journal of Mathematics, Science & Technology Education, 3*(4), 351-361.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education, 16*(2). Retrieved November 10, 2013, from http://www.amstat.org/publications/jse/v16n2/zieffler.html

APPENDICES

Appendix A

Permission to Conduct the Research

<u>ETSU IRB Final Approval</u>

IRB APPROVAL – Initial Exempt

October 9, 2014

James Ramey

RE: Examining the Relationship between the Combination of Short Online Videos and Formative Quizzes and the Statistical Reasoning Abilities of Students in an Undergraduate Statistics Course
IRB#: c0914.9e
ORSPA#: n/a

On **September 25, 2014**, an exempt approval was granted in accordance with 45 CFR
46. 101(b)(1). It is understood this project will be conducted in full accordance with all applicable sections of the IRB Policies. No continuing review is required. The exempt approval will be reported to the convened board on the next agenda.

- xform New Protocol Submission; Northeast State Approval Pending Email; Northeast State IRB Final Approval Letter (dated 10/1/14); References; CV; Protocol Sheet

Projects involving Mountain States Health Alliance must also be approved by MSHA following IRB approval prior to initiating the study.

Unanticipated Problems Involving Risks to Subjects or Others must be reported to the IRB (and VA R&D if applicable) within 10 working days.

Proposed changes in approved research cannot be initiated without IRB review and approval. The only exception to this rule is that a change can be made prior to IRB approval when necessary to eliminate apparent immediate hazards to the research subjects [21 CFR 56.108 (a)(4)]. In such a case, the IRB must be promptly informed of the change following its implementation (within 10 working days) on Form 109 (www.etsu.edu/irb). The IRB will review the change to determine that it is consistent with ensuring the subject's continued welfare.

Sincerely,
Stacey Williams, Chair
ETSU Campus IRB

Approval email from Northeast State Community College

Ramey, James M

From:	Church, Connie R
Sent:	Tuesday, August 12, 2014 3:45 PM
To:	Ramey, James M
Subject:	IRB Request - Pending Approval

Congratulations J. Mike! The Northeast State Community College Institutional Review Board has voted to grant pending approval for your research request. Full approval shall be granted following your submission of your IRB approval from ETSU.

Smiles, Connie

Connie Church
Director of Research and External Reporting
Office of Institutional Effectiveness
Northeast State Community College
CRChurch@NortheastState.edu
Phone 423.323.3191 ext. 3478

Appendix B

Demographic Survey

First Name (as it appears on your student record) (please print):

____

Last Name (as it appears on your student record)(please print):

____

Age: __________

Gender: ____________

To access the assessment, please navigate to

https://apps3.cehd.umn.edu/artist/user/scale_select.html

and enter the following access code:

****

Appendix C

ARTIST Scale for Sampling Variability

ARTIST Scale
Sampling Variability

Developed by the Web ARTIST Project
https://app.gen.umn.edu/artist/

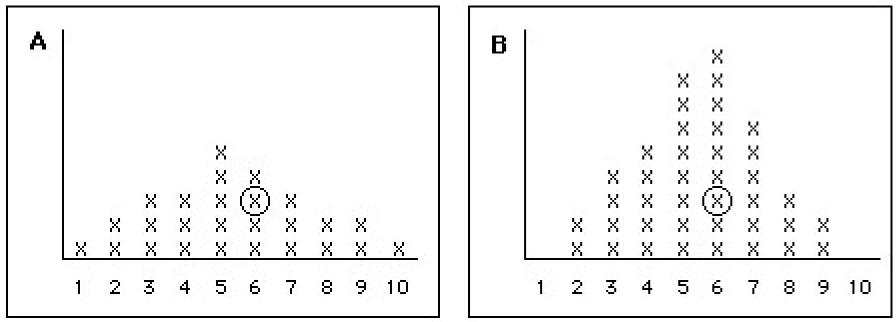
Funded by a grant from the National Science Foundation
NSF CCLI ASA- 0206571

Principal Investigators:
Joan Garfield and Bob delMas, University of Minnesota
Beth Chance, Cal Poly – San Luis Obispo
Post-doctoral Research Assistant: Ann
Ooms, University of Minnesota

April, 2006

ARTIST SCALE: **SAMPLING VARIABILITY**

1. Figure A represents the weights for a sample of 26 pebbles, each weighed to the nearest gram. Figure B represents the mean weights of a random sample of 3 pebbles each, with the mean weights rounded to the nearest gram. One value is circled in each distribution. Is there a difference between what is represented by the X circled in A and the X circled in B? Please select the best answer from the list below.



a. No, in both Figure A and Figure B, the X represents one pebble that weights 6 grams.

b. Yes, Figure A has a larger range of values than Figure B.

C. Yes, the X in Figure A is the weight for a single pebble, while the X in Figure B represents the average weight of 3 pebbles.

2. In a geology course, students were learning to use a balance scale to make accurate weighings of rock samples. One student plans to weigh a rock 20 times and then calculate the average of the 20 measurements to estimate her rock's true weight. A second student plans to weigh a rock 5 times and calculate the average of the 5 measurements to estimate his rock's true weight. Which student is more likely to come the closest to the true weight of the rock he or she is weighing?

A. The student who weighed the rock 20 times. b.

The student who weighed the rock 5 times.

c. Both averages would be equally close to the true weight.

3. Suppose half of all newborns are girls and half are boys. Hospital A, a large city hospital, records an average of 50 births a day. Hospital B, a small, rural hospital, records an average of 10 births a day. On a particular day, which hospital is less likely to record 80% or more female births?

   A. Hospital A (with 50 births a day), because the more births you see, the closer the proportions will be to .5.

   b. Hospital B (with 10 births a day), because with fewer births there will be less variability.

   c. The two hospitals are equally likely to record such an event, because the probability of a boy does not depend on the number of births.

4. A random sample of 25 college statistics textbook prices is obtained and the mean price is computed. To determine the probability of finding a more extreme mean than the one obtained from this random sample, you would need to refer to:

   a. the population distribution of all college statistics textbook prices.

   b. the distribution of prices for this sample of college statistics textbooks.

   C. the sampling distribution of textbook prices for all samples of 25 textbooks from this population.

5. Consider the distribution of average number of hours that college students spend sleeping each weeknight. This distribution is very skewed to the right, with a mean of 5 and a standard deviation of 1. A researcher plans to take a simple random sample of 18 college students. If we were to imagine that we could take all possible random samples of size 18 from the population of college students, the sampling distribution of average number of hours spent sleeping will have a shape that is

   a. Exactly normal.

   B. Less skewed than the population.

   c. Just like the population (i.e., very skewed to the right).

   d. It's impossible to predict the shape of the sampling distribution.

6. Imagine you have a huge jar of candies that are a generic version of M&Ms. We know that 40% of the candies in the jar are brown. Imagine that you create a sample by randomly pulling 20 candies out of the jar. If you repeated this 10 times to create 10 samples, each with 20 candies, about how many browns would you expect to find in each of the 10 samples?

   a. Each sample would have exactly 8 brown candies.

   b. Most of the samples would have 0 to 8 brown candies.

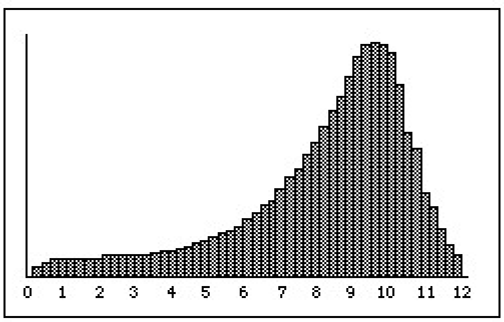
   c. Most of the samples would have 8 to 20 brown candies.

   D. Most of the samples would have 6 to 10 brown candies.

   e. You are just as likely to get any count of brown candies between 0 and 20.

Items 7 and 8 refer to the following situation:

The distribution for a population of measurements is presented below.



7. A sample of 10 randomly selected values will be taken from the population and the sample mean will be calculated. Which of the following intervals is MOST likely to include the sample mean?

   a. 4 to 6

   B. 7 to 9

   c. 10 to 12

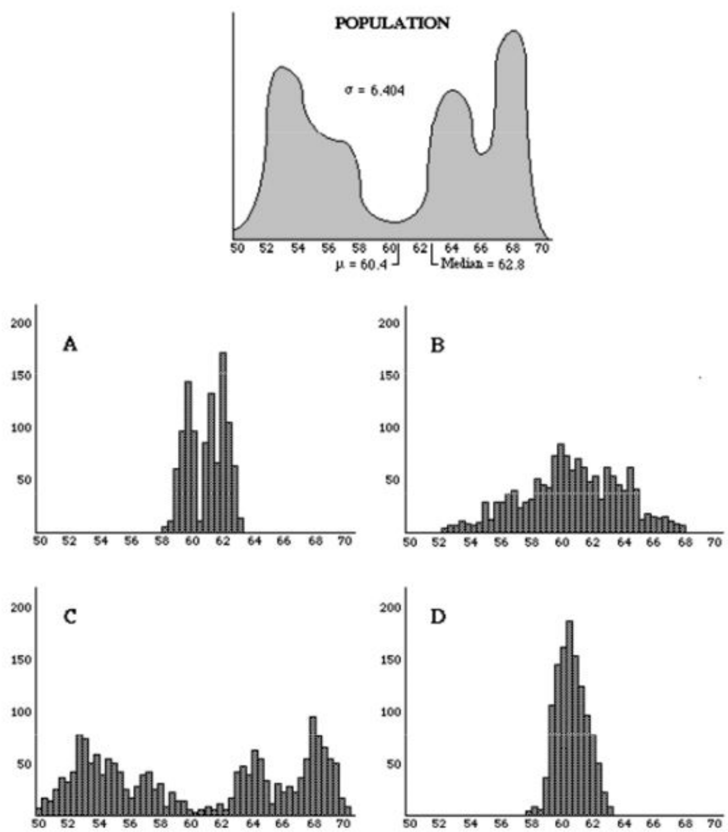
8. Another sample of 10 randomly selected values will be taken from the population and the sample mean will be calculated. Which of the following intervals is LEAST likely to include the sample mean?

   A. 0 to 3

b. 4 to 7

c. 8 to 11

Items 9 through 14 refer to the following situation:
A hypothetical distribution for a population of test scores is displayed below. The population has a mean of 60.4, a median of 62.8, and a standard deviation of 6.404. Each of the other four graphs labeled A to D represent possible distributions of sample means for random samples drawn from the population.



9. Which graph best represents a distribution of sample means for 1000 samples of size 4?

a. A

B. B

c. C

d. D

10. What do you expect for the shape of the sampling distribution (the distribution of sample means for all possible samples of size n = 4)?

   A. Shaped more like a normal distribution then like the population distribution.

   b. Shaped more like the population distribution then like a normal distribution.

   c. Shaped like neither the population or the normal distribution.

11. What do you expect for the variability (spread) of the sampling distribution?

   a. Same as the population.

   B. Less variability than the population (a narrower distribution).

   c. More variability than the population (a wider distribution).

12. Which graph best represents a distribution of sample means for 1000 samples of size 50?

   a. A

   b. B

   c. C

   D. D

13. What do you expect for the shape of the sampling distribution (the distribution of sample means for all possible samples of size n = 50)?

A. Shaped more like a normal distribution.

b. Shaped more like the population.

c. Shaped like neither the population or the normal distribution.

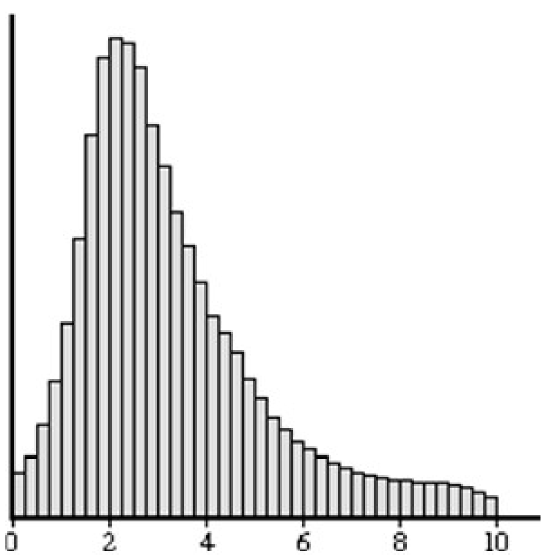
14. What do you expect for the variability (spread) of the sampling distribution?

a. Same as the population.

B. Less variability than the population (a narrower distribution).

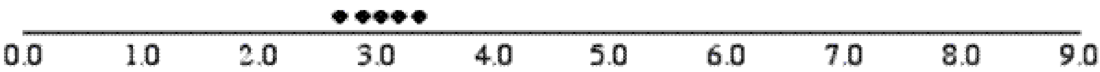
c. More variability than the population (a wider distribution).

The distribution for a population of measurements is presented below. The mean is 3.2 and the standard deviation is 2. Suppose that five students each take a sample of ten values from the population and each student calculates the sample mean for his or her ten data values. The students draw a dotplot of their five sample means on the classroom board so that they can compare them.



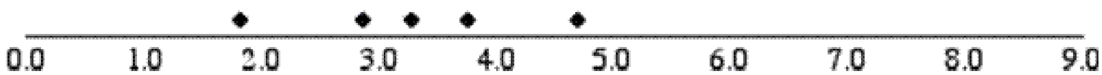
15. Which of the following dotplots do you think is the most plausible for the one they drew on the board?

a.



b.

c.



a. a.

b. b.

C. c.

VITA

JAMES MICHAEL RAMEY

Education:	B.A. History, East Tennessee State University, Johnson City, Tennessee 1993
	M.A. History, East Tennessee State University, Johnson City, Tennessee 1995
	Ed.D. Educational Leadership, East Tennessee State University, Johnson City, Tennessee 2015
Professional Experience:	Distance Education Coordinator, Northeast State Community College; Blountville, Tennessee, 2007-2015
	Director, Distance Education Programs and Services, Northeast State Community College; Blountville, Tennessee, 2015
Conference Presentations:	Ramey, J. M. (2014, May). *Banishing misconceptions: The confluence of multimedia learning, formative assessment, and conceptual change.* Canadian Network for Innovation in Education. Kamloops, BC, Canada.
Honors and Awards:	Northeast State Targeted Leadership Academy
	Tennessee Alliance for Continuing Higher Education Leadership Academy